(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number
WO 01/94599 A1

(51) International Patent Classification[7]: C12N 15/63, A61K 39/08

(21) International Application Number: PCT/SE01/01280

(22) International Filing Date: 7 June 2001 (07.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0002139-4 7 June 2000 (07.06.2000) SE
0101479-4 26 April 2001 (26.04.2001) SE

(71) Applicant *(for all designated States except US)*: **SMITTSKYDDSINSTITUTET** [SE/SE]; S-171 82 Solna (SE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BURMAN, Lars, G.** [SE/SE]; Södermannagatan 64A, S-116 65 Stockholm (SE). **ÅKERLUND, Thomas** [SE/SE]; Salagatan 15A, S-753 30 Uppsala (SE). **MUKHERJEE, Kakoli** [IN/IN]; 324B, B.B. Chatterjee Road, Kasba, Calcutta 700 042 (IN). **KATAGIHALLIMATH, Nainesh** [IN/IN]; 57, IInd Main, III stage, IIIrd Block, Basaweshwarnagar, Bangalore 560 079 (IN).

(74) Agents: **NILSSON, Brita** et al.; Stockholms Patentbyrå Zacco AB, P.O. Box 23101, S-104 35 Stockholm (SE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/94599 A1

(54) Title: GENE EXPRESSION CASSETTE AND ITS USE

(57) Abstract: A gene expression cassette comprising a secretory leader sequence encoding a signal peptide from *Clostridium difficile* having an amino acid sequence selected from SEQ ID NO: 1-12 and signal peptides of analogous exported clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked to a DNA sequence encoding a heterologous polypeptide, and optionally a part of SEQ ID NO: 22-233, is described. The gene expression cassette is inserted into a vector, and the vector is used to transformed a host organism. Compositions, formulations, vaccines and medicaments based on spores of such engineered host organisms are used e.g. for colonization of a mammal.

Gene expression cassette and its use

The present invention relates to a gene expression cassette and in particular to the use of the cassette in methods for presenting polypeptides on the surface of bacterial cells and/or secreting them into the surroundings of the latter. The invention further relates to gene expression constructs that are used to transform bacterial host cells. Uses of the invention include immunisation, in particular mucosal immunisation, induction of immunological tolerance and anti-tumour therapy in humans and animals. The intended vaccines and anti-cancer agents will also make use of bacterial spores produced by clostridia, *e.g. Clostridium difficile,* for both industrial production of the vaccine and for local production of the desired polypeptides at the body sites desired.

Vaccines against infection represent the greatest advance in medicine with unparalleled impact on morbidity and mortality at relatively low cost. Despite their cost-effectiveness, the cost associated with modern vaccines is still of concern and limits their use, particularly in developing countries. A number of factors contribute to the cost of injectable vaccines including the requirements for vaccines with defined sub-cellular components, for purity and sterility of the vaccine preparations, for testing of administration routes and combinations with different adjuvants, for maintaining the cold chain in distribution of the vaccines, and for using sterile syringes and needles. The need for repeated vaccinations also contributes to increased costs. Furthermore, for many infectious diseases a vaccine has not yet been made available

Also, in recent years great interest has been shown in mucosal immunisation, i.e. the exposure of mucosal surfaces to an antigen to elicit a general humoral and mucosal immune response, i.e. also at distant sites (Mucosal Immunology, Ed. P.L.Ogra et al., Acad. Press, 1999). Although this field is in its infancy, two such products are so far on the market, namely the oral polio vaccine and an oral (drinkable) vaccine against cholera and diarrhea due to *Escherichia coli.* The latter is an inactivated vaccine containing killed *Vibrio cholerae* organisms plus the cholera toxin subunit B which is non-toxic, immunogenic and shared between the cholera toxin and the toxin of "enterotoxigenic *E. coli*" (ETEC), the main cause of "travellers diarrhea". It is hoped that mucosal immunisation via for example intranasal or peroral administration of the antigen will provide a real alternative to injectable vaccines. As

most microbes invade humans and animals *via* mucous membranes we anticipate that the mucosal route will turn out to be the superior alternative for vaccination in many instances.

The scientific consensus at this point appears to be that live vaccines are potentially superior candidates for single-dose long lasting vaccination, because the carrier organism will continue to produce the antigen and boost immunity *in vivo*. However, experience with the bacterial carriers currently studied, for example the intestinal bacteria *Salmonella* and *E. coli* is that they are not generally classified as safe and are further difficult to distribute. Therefore, probiotic organisms such as *Lactobacilli* have been proposed as suitable carriers of foreign antigens. Problems have been identified, these include limited shelf-life unless freeze-dried which reduces viability, and difficulties in colonising the gut of the human recipient and evoking an immune response.

A classical way of enhancing the immunogenicity of vaccines is co-administration of so-called adjuvants. These may be molecules such as aluminium hydroxide or lipid vesicles that increase the exposure time for the vaccine by slowing its removal from the site of injection or "danger molecules" of microbial origin that increase the immune response in a non-specific way. Thus, in recent years it has been found that adjuvants also act by evoking production of immunomodulatory peptides called cytokines and chemokines (Brewer JM, Alexander J, Cytokines Cell Mol Ther 4:233-246, 1997. Ulanova M, Classical and non-classical antigen-presenting molecules in immune responsiveness, Thesis, Göteborg University, Sweden, 2000, ISBN 91-628-4228-5). Due to possible adverse effects the use of cytokines themselves as adjuvants must await a better understanding of optimal selection and dosing of these molecules together with vaccines. It is likely that applying cytokine adjuvants by the mucosal route is less critical than parenteral administration from a side effect point of view.

Interest has also been shown in the extracellular presentation of foreign protein antigens by bacteria, included as part of bacterial surface layer proteins. A surface layer (S-layer) protein is herein defined as any molecule of proteinaceous nature, including e.g. protein, glyco- or lipoprotein occurring in the outer layer of a bacterium and capable of being exposed on the surface of the bacterium. S-layer proteins are a main constituent of the cell wall of some gram-positive bacterial genera. They may be continuously and spontaneously produced in larger amounts than any other class of protein in the cell. WO-95/19371 describes a fusion protein of at least a part of a S- layer protein and a heterologous peptide, the intention being that the polypeptide is expressed and presented on the surface of the cell. A range of

bacterial hosts is mentioned including *Staphylococcus*, *Streptococcus*, *Bacillus*, *Clostridium* and *Listeria*. A preference for *Bacillus* is stated and the examples use *B. sphaericus*. Elsewhere, WO-97/28263 describes processes for the recombinant preparation of S-layer proteins in gram-negative host cells. It is suggested that these proteins could include antigenic species. FR-A-2778922 describes the use of genes which regulate the synthesis of toxin products in *Clostridium* bacteria, to produce polypeptides.

We first set out to investigate toxin release and to identify other extracellular proteins produced by *Clostridium difficile* (*C. difficile*). *C. difficile* is an anaerobic spore forming pathogen causing *C. difficile* associated diarrhea (CDAD) and pseudomembraneous colitis (PMC) by producing two toxins, A and B.

In contrast to studies by Kamiya et al (J.Med.Microbiol.,1992, 37, 206-210) and Ketley et al (J.Med.Microbiol.,1984, 18, 385-391) we found that the accumulation of extracellular toxins is not accompanied by cell lysis suggesting a toxin export mechanism.

Our first detailed analysis (described in Example 9 below) of proteins occurring extracellularly as well as in the cell wall and membrane fraction of strains VPI 10643 and 630 and analysis of the DNA/genes encoding these proteins revealed a genomic segment containing eighteen genes (Figure 1A). Seven proteins (ORF1,3,5-7,9 and 11), when compared with publicly available sequence showed some homology to N-acetyl muramoyl L-alanine amidase (CwlB/LytC) and modifier protein of major autolysin (LytB) from *B. subtilis*, as well as S- layer proteins from *Lactobacillus spp* (Tables 1 and 4). The amidase motif was located either at the C-terminal or the N-terminal end of the S-layer protein ORFs (see examples in Figure 1B). Other ORFs showed similarity to genes involved in polypeptide secretion (ORF2/*secA*), polysaccharide and capsule synthesis, and possibly glucosylation of the S-layer proteins. Further database searches indicated that the amidase motif confers anchorage of the S-layer proteins to the clostridial cell wall peptidoglycan-teichoic acid.

A search in the revised *C. difficile* database revealed five additional genes upstream of ORF1 which had similarities to the previously found ones, i.e. they had a two-domain architecture one showing homology to the CwlB/LytC and LytB proteins. These ORFs thus had the putative cell wall binding amidase motif typical of the other S-layer ORFs and were designated D, E, G, H and I by us (Fig. 2 and Table 1).

Significantly, we found that the N-termini of all S-layer ORFs contained a typical signal peptide for export via *sec*-dependent secretion and that for e.g. ORF1 the predicted signal peptide cleavage site (Figure 3) was identical to that found in the protein sequence .

Furthermore, the secreted ORF1 product was further cleaved into two peptides in strain 630, the C-terminal one containing the N-acetyl muramoyl L-alanine amidase like sequence (Figure 1B and Example 9). Although we identified the cleavage point, the kinetics and precise mechanism of this proteolytic event remains unknown. It is likely that the S-layer protein cleavage product containing the amidase motif provides cell-wall binding, whereas the other peptide showing more sequence variability is more surface exposed and providing antigen variation between strains (serotypes). We found no significant match between sequences within ORF1 and the S-layer homology motif (SLH domain) found in most presently known S-layer proteins although a weak similarity to an S-layer protein from *Lactobacillus helveticus* was found for the N-terminal part of ORF1 (Figure 1B). Also, apart from the shared amidase motif, we found no significant homology beetween the different S-layer ORFs that could suggest how the variable cleavage product(s) get anchored to the cell-wall binding peptide (inner S-layer) or to each other to form the outermost layer recently distinguished by electron microscopy (Cerquetti M et al. Characterization of surface layer proteins from different *C. difficile* clinical isolates. Microb Pathogenesis 28: 363-372, 2000).

Our subsequent analysis of 21 *C. difficile* serogroup type strains, probably representing all major genetic lineages of the species, indicated that a gene segment corresponding to ORF1 and its upstream region plus ORF 2 is generally present in *C. difficile* (Example 1C).

The N-terminal part of ORF6 showed homology to eukaryotic cysteine proteases (Fig. 1B). ORF5 has been suggested to be involved in adhesion to epithelial cells (Abstract; The Third International Meeting on the Molecular Genetics and Pathogenesis of the Clostridia, June 8-11, 2000, Chiba, Japan).

The present invention is based, at least in part, on the above discoveries. We have identified and developed a polypeptide expression and secretion system that may be used to produce a desired polypeptide on the surface of and/or into the surroundings of bacteria, for introduction into an appropriate mammal. The system may be used for example to initiate mucosal vaccination. A particular advantage of the system is that it may be used with any convenient *Clostridium* species, independently of any normal S-layer protein production. Furthermore, in case of *C. difficile* it is possible to use strains lacking the 5-gene toxicity cassette encoding the two major virulence factors toxins A and B and thus avoid the risk of CDAD when administering the engineered peptide producing strains to humans aged 2-4 years or more (neonates and small children are insensitive to the toxins) or animals.

Therefore, in a first aspect of the invention we provide a gene expression cassette comprising a secretory leader sequence selected from any one of ORF1, ORF3, ORF5-7, ORF9 or ORF11 (SEQ ID NO: 1 –7) (c.f. Figure 1 and Table 1) of *C. difficile* strain 630 linked to a DNA sequence encoding a heterologous polypeptide. Alternatively, the secretory leader sequence is from any one of ORF D, E, G, H and I (SEQ ID NO: 8 – 12) (cf. Figure 2 and Table 1) or from any analogous S-layer ORF taken from any *C. difficile* strain.

By "heterologous" we mean a nucleic acid sequence or protein not native to the clostridial strain being used.

Use of each of the secretory leader sequences mentioned above represents a separate and independent aspect of the invention. The secretory leader sequence is preferably from ORF1.

A recent publication by Karjalainen et al, Infection and Immunity, May 2001, p3442-3446 provides confirmation and analysis of most of the ORFD - ORFI - ORF1 - ORF11 S-layer gene cluster in *C. difficile* strain 630. The nucleotide and polypeptide sequences disclosed by Karjalainen et al are incorporated herein by reference.

In a further aspect of the invention the gene expression cassette further includes a promoter of prokaryotic origin. The promoter is preferably a strong promoter and in general is placed 5' of the secretory leader sequence in the gene cassette.

In a further aspect of the invention the gene expression cassette further includes a DNA sequence encoding at least a functional portion of an S-layer protein of *C. difficile* fused to a nucleic acid coding sequence coding for a heterologous polypeptide such that the resulting fusion polypeptide will be expressed and presented on the outer surface of the host cell harbouring the cassette. If desired the polypeptide can also be released from the bacteria, e g. by excluding the S-layer amidase motif from the construct (cf. Figure. 3).

In a further aspect of the invention the engineered gene expression cassette optionally further comprises at least a functional part of the secretory (*secA*) gene represented by ORF2. This may be used to complement or replace the function of the normal *C. difficile sec* gene in order to ensure efficient translocation of the peptide(s) produced by the cassette across the cytoplasmic membrane.

An example of a preferred gene expression cassette is conveniently illustrated in Figure 3.

The promoter in the gene expression cassette is conveniently a strong promoter, this may be the native promoter for ORFs 1 - 12 of *C. difficile* of strain 630 (Table 1).

Alternatively, the promoter sequence is from any one of ORF D - I (cf. Table 1), alternatively from any other analogous S-layer ORF from a *C. difficile* strain or from another gene, preferably from this species (see Specific description B1). The promoter may thus be another prokaryotic promoter that is strong, inducible or constitutive, and functional in the polypeptide producing bacterium. In all potential applications a distinct advantage of this cassette is the very large amounts of protein produced and exported.

The gene expression cassette is conveniently placed in a vector or specifically a plasmid carrying a transposon belonging to for example the Tn916, Tn5387 or the Tn5398 families. After transfection of a *C. difficile* host organism these transposons are able to insert themselves into its chromosome thereby making the engineered cassette a stable trait of the bacterium (cf. Figure 4). For other *Clostridia* other vectors may be preferable, e.g. the engineered shuttle plasmid pJIR750. Unlike the *C. difficile* plasmids currently available, this vector can replicate within both an *E. coli* and a *C. perfringens* host and is not dependent on integration of the plasmid into the host chromosome. Any convenient *Clostridium* species may be used, to date over 70 species have been defined by rRNA sequence analysis. These include *C. difficile* and classical pathogens as *C. perfringens, C. tetani* and *C. botulinum,* also *C. acetobutylicum* that is being genetically manipulated and used for industrial production of acetic acid and *C. beijerinckii* that has been transformed with *E. coli* genes.

C. perfringens is currently the species most amenable to genetic engineering. It is a normal, moderate level, fecal coloniser of most, if not all, humans. *C. difficile* is found in the fecal flora of most newborns, less often in adults but commonly in hospitalized individuals. As *C. difficile* is an early, normally colonising intestinal organism and even toxigenic strains are unable to cause CDAD in newborns and infants up to 2-4 years of age, we believe that recombinant *C. difficile* producing desired antigens and adjuvants is suitable for oral vaccination at any convenient time after birth.

Whereas *C. perfringens* normally produces many toxins about half of wild *C. difficile* strains are genetically non-toxigenic, which may be an advantage from a safety point of view. *C. difficile* toxin negative strains are preferred as host cells for the gene expression cassettes of this invention, at least for individuals aged 2-4 years or more (see above).

The nucleic acid sequence coding for a heterologous polypeptide is placed in the gene expression cassette before or after insertion into a convenient vector or plasmid. The insertion points for the nucleic acid sequence are at the discretion of the skilled scientist, there may be in the variable or in the constant region of the relevant ORF nucleotide sequence. Routine

experimentation may be used to determine convenient and particular insertion points. In Figure 3 we disclose polypeptide cleavage sites that need to be taken into consideration (See Specific description B3).

Examples of convenient plasmids include those mentioned in Figure 4, for example pCI195 and pSMB47. Convenient transposons include those belonging to the Tn916, Tn 5397 and Tn5398 families for transfection into *C. difficile* and for example pJIR750 for *C. perfingens* or other *Clostridia*. Any convenient heterologous nucleic acid sequence may be placed into the gene expression cassette. In a further aspect of the invention we provide a vector or plasmid comprising a gene cassette of the invention.

The vector or plasmid may then be transfected into a convenient host using techniques known in the art (see for example: Gene 82: 327-333, 1989). For *C. difficile* it is at present preferred to introduce the plasmid into a *Bacillus* species such as *B. subtilis* and then transfer the target DNA by filter mating (conjugation) into a convenient *C. difficile* strain (outlined in Fig. 4). This will generally require the use of a conjugative transposon-bearing plasmid such as pCI195 or pSMB47 (J. Antimicrob. Chemother. 35: 305-315, 1995; FEMS Microbiol. Lett., 168: 259-268, 1998; D. Lyras, J. I. Rood, Clostridial genetics, in Gram-positive pathogens, ed. V. A. Fischetti, Am. Soc. Microbiol, 2000). However, we anticipate that further materials and procedures will become available for the direct introduction of plasmids or other foreign DNAs into *Clostridia* and particularly *C. difficile*. For e.g. *C. perfringens* such vectors and techniques are to some extent already available.

In a further aspect of the invention we provide a Clostridial bacterium transformed with a gene expression cassette of the invention encoding the desired fusion peptide or entirely heterologous polypeptide(s).

The transformed Clostridial bacterium, when administered orally to any convenient mammal such as a human or animal will lead to the intestinal colonization, production and presentation of the desired polypeptide particularly in the large bowel that is the natural site of colonization of *C. difficile*. The bowel wall is surrounded by an immense immune apparatus, the so-called Peyer's patches and thus, specialized in mounting immune responses of various types. Large bowel colonization by a clostridial vaccine or peptide producer strain thus enables a much longer immune stimulus than a traditional injection. In contrast to clostridia, the alternative and much studied S-layer producers for vaccine purposes, *Bacillus spp*, are free-living, obligate aerobic bacteria and unable to replicate in the anaerobic bowel lumen and

thus, unable to colonize a recipient mammal. For clostridial colonization and peptide delivery in hypoxic tissues iv administration is used.

It will be appreciated that Clostridia carrying the gene expression cassette of this invention including DNA encoding different heterologous peptides allows the highly efficient production and export of these polypeptides in hypoxic tissues after iv administration, or into the gut, particularly the colon, of the orally colonized individual for a variety of prophylactic or therapeutic uses. Another advantage of this gene cassette for expression of heterologous peptides is its versatility, i.e. that it is normally used to produce and export peptides of varying size and having completely different amino acid sequences, in their N-terminal or C-terminal end.

In further independent aspects of the invention the recombinant gene expression cassette is used to produce in the gut, for example

(i) peptides and enzymes for therapy and prophylaxis of various diseases, e.g. peptides having specific antimicrobial activity, cytokines against inflammatory bowel disease, and β-lactamases to prevent diarrhea due to antibiotic therapy

(ii) single, fusion or multiple polypeptide antigens of microbial, animal or mammalian origin for neonatal immune balancing, vaccination against infections, allergy, metabolic or auto-immune disease, cancer, (in)fertility, and drug addiction.

(iii) carrier molecules (so-called adjuvants) separate or fused to the antigen in order to amplify or modulate the immune response to the antigen in a desired way according to (ii), e.g. a strong IgA response against a mucosal invader.

In a still further aspect of the invention the gene expression cassette of the invention may be used to provide recombinant clostridia for local production of peptides in tissues after iv administration of their spores (see below), for example for the prophylaxis and/or treatment of fibrinolysis in arterial or venous occlusion and/or for revitalising gangrenous and/or necrotic tissue in various diseases. Furthermore, for anti-tumour therapy by local production of

(i) immune stimulating human peptides for improving tumour host defence,

(ii) enzymes that convert a pro-drug to a cytostatic agent inside a tumour (thus avoiding systemic side effects)

(iii) cytotoxins of e.g. bacterial origin to destroy tumour cells

(iv) angiogenesis inhibitors at local concentrations enough to prevent local blood vessel formation and thus, tumour growth

(v) signal transduction inhibitors.

In a further aspect of the invention we provide a pharmaceutical or veterinary composition which comprises a transformed viable *Clostridial* cell with the ability to present and/or to secrete the desired polypeptide together with a pharmaceutically or veterinary acceptable carrier or diluent.

The composition may be formulated as a vaccine. The composition may be administered orally, or intranasally or alternatively, the polypeptide can be isolated, purified and administered parenterally, e.g subcutaneously or intramuscularly.

The amount of the desired peptide(s) presented and/or secreted by the transformed strain may be modulated in the body by using

(i) promoters with different strength (power),

(ii) a promoter or regulator responding to external stimuli (inducible, e.g. by a specific carbohydrate) normally present in the gut or administered together with the engineered bacterium,

(iii) different dosage regimens (number of bacteria per dose and doses per time period) or

(iv) methods that influence the ability of the strain to colonise and propagate in the gut for convenient periods of time. Relevant factors include ability to compete with other bacteria, adhere to mucosal cells, and to avoid expulsion by local immune response mechanisms. The latter can be achieved by exploiting induction of tolerance to the natural S-layer antigen of the strain in the neonate (see above and below) or by using its putative normal antigenic variation, suggested to us by the presence of the large number of S-layer ORFs present in *C. difficile*, either by allowing the vaccine strain to change serotype during natural long-term colonization or by repeated applications (colonizations) over time of different strains producing the same antigen but having a different serotype antigen.

The transformed *Clostridia* as anerobic organisms are conveniently produced by fermentation under for example low oxygen tension and purified and recovered as known in the art for native *Clostridia*, for example by washing and freeze-drying. They may be formulated together with excipients as needed, for example magnesium stearate, lactose, or carboxymethyl cellulose, into solid dosage forms, e.g. in capsules, predominantly for oral administration. The dosage forms may be protected against the acidity of the stomach by a

suitable enteric coating, comprising for example Eudragite "S", Eudragite "L", cellulose acetate, cellulose phthalate or hydroxypropyl cellulose. A preferred dosage form comprises freeze-dried transformed *Clostridia* contained in vials or ampoules, optionally under inert gas. Preferably, the transformed *Clostridia* cells are administered orally or intranasally, as an aqueous, reconstituted suspension of the lyophilized cells e.g in water or physiological saline, optionally with addition of pharmaceutically acceptable buffers, e.g. sodium bicarbonate, phosphate or citrate to keep the pH of the suspension between 6 and 8, preferably between 6.5 and 7.5.

The dosage forms produced as described above may comprise a mixture of viable and non-viable bacteria depending on the process and/or the storage conditions. The viable, transformed *Clostridia* will, after oral administration, become attached to those parts of the gut, for example the lower intestinal tract, which provide appropriate growing conditions and proliferate, producing the desired polypeptide in increasing amounts. This will provide for an enhanced and sustained physiological effect, for example immunisation, of the polypeptide.

If exposure to defined amounts of the polypeptide is desired, non-viable transformed *Clostridia* presenting the polypeptide can be administered. The non-viable cells can be obtained as known in the art, e.g. by exposing the live cells to agents, e.g. heat, formaldehyde, antibiotics or solvents, which kill them. It is also possible to use cell walls (sacculi) or to use S-layer fragments obtained by mechanical or other disruption of the bacterial cells. These agents can be formulated into pharmaceutical and veterinary compositions as described above for live transformed *Clostridia*.

By "secretion" or "release" we mean that the heterologous polypeptide is exported out from the host cell into the surrounding environment as a soluble antigen. This is conveniently achieved by fusing the DNA coding for the polypeptide to a DNA sequence coding for a signal peptide sequence of any one of ORF1,ORF3, ORF5-7, ORF 9 or ORF11 (SEQ ID NO:1-7, cf. Figure 1 and Table1) preferably to that of ORF1 and expressing it as described above under control of a strong promoter and exporting it with the aid of the *sec* gene (ORF2) product. Alternatively, the DNA codes for a signal peptide sequence of any one of ORF D, E, G, H or I (SEQ ID NO: 8-12, cf. Figure 2 and Table 1), or that of any other suitable secreted bacterial protein.

By "presentation" we mean that the polypeptide is translocated across the cell membrane and presented on the surface of the bacterium in a sufficient manner for it to act as, for example, a particulate antigen. The DNA coding for the heterologous polypeptide may

then for example be fused to a S-layer coding sequence, which codes at least for a functional cell wall binding portion of a S-layer protein of *C. difficile* (Figure 3) and expressed as described above to get exposure of the heterologous polypeptide on the outside of the host cell and thus, hooked to the amidase motif of the S-layer protein. Alternatively, omitting this motif from the construct in order to get increased release of the heterologous peptide (see Specific description B3).

The heterologous polypeptide may be a foreign epitope or immunogen giving rise to antibodies that protect against disease, we note that many antibodies elicited are not protective. It typically comprises an antigenic determinant of a pathogen. The pathogen may be a virus, bacterium, fungus, yeast or parasite. The antigen may also be a "self" molecule for prevention or cure of disease (see below). The heterologous polypeptide may further be an antimicrobial peptide, e.g. for elimination of undesired microorganisms, and an anti-tumour peptide (see below) or a molecule that changes the immune response of the gut from a negative one, such as allergy or auto-immune tissue destruction, to a positive one, such as infection protection (see above). For example, *Lactobacillus* components are believed to prevent allergy development and live lactobacilli are currently given to infants in successful trials for prevention of allergy (Björkstén B, pers comm, and Kalliomäki et al. Lancet 357: 1076-1097, 2001). Cystein proteases such as cathepsin are thought to change the intestinal mucosal response to infection from a Th2 type (disease promoting) to a Th1 response (infection protection). Alternatively, the polypeptide(s) may be enzyme(s) that improve digestion of food, or that together synthesize a polysaccharide antigen of a microorganism, an antibiotic, or a specific vitamin or other nutrient or hormone useful to the host mammal. An enzyme produced by the the engineered *Clostridial* bacterium may also be an antibiotic inactivating enzyme, e.g. a beta-lactamase, to be given together with or after the antibiotic for prevention of CDAD or non-specific antibiotic induced diarrhoea, common problems in hospitals today.

The heterologous polypeptide may also be a part of an antibody molecule. This may comprise the constant part in order for example to obtain an enhanced non-specific immune response or the response to a co-administered antigen (adjuvant effect). Alternatively, it may be the variable part directed against any surface or secreted component of a microorganism (toxin, antigen, adhesin) in order to prevent its ability to colonize and cause intestinal disease. The expression product of the cassette of the invention may also represent the immune

stimulating part of allergy causing antigens lacking their IgE interacting part, thus evoking an antibody response but avoiding an allergic reaction (anti-allergy vaccination).

Whether the heterologous polypeptide is to be provided alone or fused with a carrier peptide, or presented cell-bound, released or both depends on its desired function. For example, for a polypeptide acting as an enzyme, free "secreted" molecules may be most effective, whereas in case of vaccination an antigen fused to a carrier peptide or being a cell-bound ("presented") polypeptide on a bacterium, strongly adhering to or being phagocytosed by the gut mucosa, may give the best mucosal immune response.

The immune response to a heterologous peptide may be increased by fusion to the repeating C-terminal sequences encoding the non-toxic motifs of the *C. difficile* toxins A and B that enable these to enter the colonic mucosal cells by receptor-mediated endocytosis, and/or to a portion of toxin B responsible for intracellular and intercellular spread of the antigen (see Barth et al below). Thus, by using adhering clostridial bacteria producing a desired heterologous peptide antigen fused to non-toxic parts of the *C. difficile* toxins, the mucosal immune reponse may be boosted (adjuvant effect).

A further improved immune response may be obtained by exploting the natural S-layer proteins of *C. difficile* that seems to anchor the organsim to the mucosa. It is likely that the amidase like fragments are directed inwards to provide cell wall anchorage, whereas the sequence unique fragments represent the the outermost portion of the S-layer protein serving as surface antigen (see above) and probably also as adhesin by which *C. difficile* attaches to the mucosal cell surface as recently suggested by Waligora et al. (Infect Imm 69:2144-2153, 2001). Thus, by switching between expression of its different S-layer ORFs over time each *C. difficile* strain may achieve surface antigen variation and thus, immune evasion and prolonged colonization in the gut. However, after deliberate colonization of a newborn with a certain *C.difficile* strain, particularly if expressing or being administered together with a "danger molecule", tolerance (see above and below) to its S-layer serotype protein may be obtained enabling the use of the same *C. difficile* serotype for efficient deliberate long-term colonization of this individual, e.g. for vaccination purposes, also during later periods of life.

In a further aspect a carrier peptide or adjuvant, e.g. a "danger molecule" is used in addition to the desired heterologous polypeptide, administered or produced *in vivo* either as a separate molecule or fused to the principal (antigenic) polypeptide. This is in order to amplify desired specific immune responses for prevention or therapy of infection, or in the neonate also for shifting its general response towards anti-infection and tolerance of "self" and

therefore away from allergy and auto-immunity (see above and below). The "danger molecule" or adjuvant is a species that may stay in a human or animal body for a long time, such as up to one, three, six months or up to one year. Alternatively, or in addition, this species is capable of eliciting a stronger immune response that the desired heterologous polypeptide acting alone. "Danger molecules" are often of microbial origin, rapidly recognized and strongly reacted upon both by the innate/primitive and the trained/specialized immune system (see above). A convenient reference for this aspect of the invention is the thesis by Carola Rask of the Department of Medical Microbiology and Immunology at Göteborg University, Sweden, ISBN 91-628-4497-0 "Cholera toxin B subunit as a carrier for inducing mucosal immunity and/or peripheral tolerance."

By using a mixture of different genetically engineered *Clostridia* strains, presenting and/or secreting different heterologous polypeptides, which act additively or synergistically, it is possible to achieve an enhanced physiological response. Thus oral immunization with a mixture of genetically engineered *Clostridia* strains, each presenting and/or secreting a different polypeptide derived from the same or another microbial pathogen, will provide both a broader immune response and a so called adjuvant effect, i.e. a more complete immune response and a better vaccination against the pathogen than obtained by using just one strain expressing a single immunogenic epitope of the pathogen.

In a further aspect of the invention we provide a medicament or therapeutic agent which comprises a *Clostridial* bacterium transformed with a gene cassette of the invention and capable of presenting on the surface of the bacterium and/or secreting a polypeptide in a human or animal body.

The medicament or therapeutic agent is conveniently a lyophilised powder for reconstitution as a suspension or for production of a solid pharmaceutical form such as a capsule or a tablet. The therapeutic agent can be administered orally or intranasally. For oral administration capsule or tablet formulations may be used. To protect the compositions against the acidity of the stomach buffer substances, e.g. sodium bicarbonate, may be used and/or the formulations may be covered with enteric coatings, e.g. Eudragite "S" or "L", cellulose acetate, cellulose phthalate or hydroxypropyl cellulose. A convenient way for oral administration of the therapeutic agents is to provide them as lyophilised powders, and shortly before administration to suspend these in for example water, fruit juice or physiological saline, optionally with addition of sodium bicarbonate or neutral citrate, or phosphate buffer to protect against the acidity of the stomach. Any convenient dose may be used, this may be in

the range from 1 to 10^{11} bacteria, more conveniently we anticipate this to be in the range from about 10^3 to about 10^9 bacteria.

A principal use of the invention is in vaccination. Therefore in a further aspect we provide a vaccine which comprises a *Clostridial* bacterium transformed by a gene cassette of the invention and capable of secreting and/or presenting an antigen on the surface of the bacterium in a human or animal body.

In addition to improving the existing anti-infection vaccines and creating new ones, there is also current interest in several novel uses for vaccines.

Allergy One strategy is engineered anti-allergy vaccines containing the immunostimulatory part of each antigen but lacking the part which interacts with IgE and thus, normally elicits the allergic reaction. Another new approach is to induce an immune response towards human IgE, that normally governs the allergic response, by turning these molecules into "non-self" ones e.g. by coupling to IgE of animal origin. The use of these hybrid IgE molecules as vaccine is expected to elicit production of anti IgE antibodies that thus, inactivate human IgE thereby preventing allergy.

Alternatively, allergy may be prevented by stimulating the immune apparatus of the newborn in such a way that cellular, IgG and IgA antibody responses to microbial antigens, i.e. anti-infection, will be preferred to IgE production against allergens (immune balancing).

Auto-immune diseases. Another new proposed area for vaccines is to boost tolerance to "self" antigens *in utero* and/or in the newborn in order to prevent later development of auto-immune disorders such as type 1 diabetes, rheumatoid arthritis, inflammatory bowel disease and multiple sclerosis. This may be achieved either by non-specific tilting of the newborn immune system towards anti-infection and away from auto-immunity and allergy as mentioned above, or by applying the "self" molecule (e.g. human insulin or other beta cell antigens, connective or CNS tissue antigens) coupled to or together with a "danger molecule" of microbial origin (e.g. part of the tetanus or cholera toxin, see above) here in order to amplify the normal immunotolerance response to e.g. insulin and thus, the natural avoidance of juvenile diabetes.

Pregnancy and metabolic diseases. In contrast to the newborn, exposure to a "self" antigen especially when coupled to a "danger molecule" may in the adult individual lead to an immune response to the antigen rather than reinforced tolerance. Such vaccines boosting specific auto-immunity may be used for prophylaxis and therapy by eliciting antibodies directed against specific "self" target molecules, such as sperm or egg components or human

gonadotropin (hCG) to prevent fertility, enzymes in cholesterol biosynthesis to prevent arteriosclerosis, beta amyloid for prevention and cure of Alzheimer's disease, other brain proteins to counteract prion and Creutzfeld Jacobs disease.

Drug addiction. A further novel application of vaccines includes the use of drugs molecules such as nicotine or heroin as part of the antigen for induction of anti drug antibodies that block its activity and remove the drug and thereby abolishes its CNS effect, in order cure addiction.

Cancer. Novel multicomponent vaccines containing "danger molecules" may be of use also against cancer both by boosting the innate immune defense, by eliciting anti-tumour antibodies and cellular immune responses or by stimulating apoptosis of cancer cells.

Before the optimal anti-infection and other types of mucosal vaccines can be achieved a lot more has to be learned, e.g. about the interaction of *C. difficile* with M-cells and mucosal cells in the gut, the uptake, processing and presentation of each antigen, and optimizing the choice and presentation of adjuvant so that the immune response can be maximized and modulated in the desired way and thus, perfected for each application in order to obtain a desired cellular, IgA, IgG and subclass or combined response. Thus, the choice, size and form (soluble, particulate) of antigen, the vector (live?), the choice and form of carrier molecule (adjuvant, separate, fused to the antigen), recipient (mother, child, both, adult), timing of administration etc need to be tailored in each case.

In the above approaches to vaccination against infection in mammals and for immune modulation in newborns and adults recombinant clostridia producing desired popypeptides may be of particular interest as live vectors for mucosal immunization since certain species, e.g. *C. perfringens* and *C. difficile* belongs to the normal gut flora. *C. difficile* is particularly common in the neonatal period, and toxin negative strains thus can be given orally at all ages without ethical concerns. *C. difficile* appears to be a particularly good candidate also for delivery of antigens for gut mucosal immunization as exposure to microscopic numbers of the organism during hospital stay resulting in asymptomatic carriage is enough to yield an immune response to its toxins (NEJM 2000). Furthermore, we have observed in animals that asymptomatic gut colonization by *C. difficile* results in an immune response also to its S - layer protein (see below). These responses are probably enhanced by the the non-toxic part of the *C. difficile* toxins that are used for their receptor mediated pinocytosis into the mucosal cells. Toxin B then can form membrane pores in the pinocytic vacuoles containing toxin and presumably also in phagocytic vacuoles containing whole bacteria (Barth H et al, Low pH

induced formation of ion channels by *C. difficile* toxin B in target cells. J Biol Chem 276(14): 10670-10676,2001). Thereby the the toxin and other bacterial components may be released into the cytosol of the mucosal cells and may spread also to neighbouring cells including to antigen presenting cells and thus, enhancing an immune response. Such an unusual adjuvant effect of *C. difficile* toxin B obtained by breakage of phagocytic vacuoles and intercellular spread of internalized antigens and bacteria can alternatively also be obtained e.g. by including the membrane attacking peptide listeriolysin O from *Listeria monocytogenes* in a recombinant *C. difficile* strain in order to boost immunity as has been shown in experiments using other gut mucosal delivery systems (Dietrich G et al, From evil to good: a cytolysin in vaccine development, Trends in Microbiology 9:23-28, 2001).

Clostridia furthermore represent a unique torpedo able to deliver a desired heterologous polypeptide to hypoxic tissues such as tumours. This is because spores of these obligate anaerobic organisms given intravenously are known to settle and be able to germinate into growing clostridial cells in the hypoxic parts of tumours but not in healthy tissues. This phenomenon was described already in 1955 (reference 7 in Theys J et al, FEMS Immunol Med Microbiol 30:37-41, 2001) and is currently being studied for therapeutic applications by several groups (see also Abstracts, The 3rd Int Meeting on Molecular Genetics and Pathogenesis of the Clostridia, Chiba, Japan, June 8-11, 2000). Thus, anti-tumour peptides including apoptosis inducing peptides, cytokines, toxins and other proteins, such as enzymes locally converting pro-drugs to active anti-cancer chemotherapeutic agents, thus minimizing systemic side effects, all produced by recombinant clostridia inside tumours may become novel approaches to cure cancer. We propose that another approach to anti-tumour therapy is iv injection of spores of recombinant clostridia producing various angiogenesis inhibitors, mostly of peptide nature and currently used in many clinical trials (phase I: angiostatin, SU6688, combrestatin A-4 prodrug, PTK787/ZK2284; phase II: endostatin, anti-VEGF Ab, TNP-470, Il-12, 2-methoxyestradiol, squalamine, vitaxin, EMD 121974, COL-3, CGS-27023A, CAI; phase III: thalidomide, marimastat, INF-alfa, neovastat, BMS-275291, SU5416, AG3340, IM862 as summarized in Larsson H, Regulation of angiogenesis, Thesis, 2001, Uppsala University, Sweden, ISBN-91-554-4954-9).

A problem in these studies and possible to overcome by local production of enzymes, toxins and other peptides is limited effect *in vivo* due to short half-life in serum and thus, insufficient local concentrations of e.g. the cytostatic agent or the anti-angiogenesis peptide. In the latter application recombinant clostridia may become particularly powerful (self-

accelerating), because the effect of the peptide will further lower the oxygen tension and thus, enhance bacterial growth and further production of the peptide(s)

We further believe, that that iv administration of spores from recombinant clostridia may be used also against other diseases involving local tissue hypoxia such as fibrinolytic and other agents for venous or arterial occlusion, and oxygen releasing or other tissue vitalizing molecules for tissue necrosis and gangrene.

We now found and disclose that *Clostridial* spores may be used to deliver heterologous polypeptide(s) to a human or animal body. This is an important step forward. A spore is a dormant or resting state of a bacterial cell. Unlike bacterial spores from species belonging to the obligate aerobic genus *Bacillus* (see above), ingested *Clostridial* spores naturally germinate into vegetative bacteria that can grow anaerobically and naturally colonise a human or animal gut. Intake of the spores of the genetically engineered *Clostridia* is preferably through the oral route. Spores are able to resist stomach HCl and digestive enzymes. Upon contact with bile they will germinate and establish themselves in the colonic flora as vegetative bacteria presenting and/or secreting for example the desired heterologous peptide *in vivo*.

Therefore, in a further aspect of the invention we provide a therapeutic agent which comprises spores of *Clostridia* transformed with a construct capable of expressing, secreting or presenting a heterologous polypeptide in the mammalian body after conversion (germination) to live (vegetative) bacteria.

The construct is preferably a recombinant gene cassette of the invention as outlined before. The mammalian body is preferably a human or animal.

The use of *Clostridial* spores has a number of advantages including low production cost, relative ease of production, very long shelf life independently of the mode of storage, ease of administration, production of antigen at the site of action, and an oral route of immunisation which may be superior to a parenteral one. As mentioned above for bacteria, spores are suitable for administration of mixtures (coctails) of recombinant *Clostridia* having different desired properties.

The use of live vaccines administered via the oral route may lead to further fecal-oral transmission and enhanced immunization of a population. On the other hand, this may also be considered as to be unwanted spread of genetically modified organisms in the environment. Spores of *Clostridia* survive readily in the environment whereas the vegetative forms have a very limited capability to survive in an oxygen-containing milieu. The invention may be

further developed to create Clostridia that are unable to reconvert to spores, once they have germinated in the colon. One way is to modify a genetic element present in *C. difficile* that is similar to the so-called *skin* (Sigma K intervening) element of *B. subtilis.* This element truncates the sigma K factor necessary for sporulation, and becomes removed by a specific excision system during sporulation (Krogh, S. et al.(1996) and Takemaru, K. et al.(1995)). By genetic modification of the excision system, and insertion of wild-type copies with inducible promoters, there is a possibility to create a host strain that is able to sporulate only during special conditions, e.g. in the presence of a special chemical (IPTG) or at low temperature (20 °C). Such construction would allow the production of spores *in vitro*, whereas no new spores are created in the vaccinated host. The spread of genetically modified *Clostridic* microorganisms to the environment would still occur, but the probability of survival of these organisms would in practice be very low or nil.

Spores of the transformed *Clostridia* are produced, purified and isolated in the same way as for native *Clostridial* strains. They may thus be readily obtained from a stationary phase culture for example by treatment with ethanol, acid or heat or by combinations of such measures followed by purification and isolation in a conventional way. As outgrown spores will have the same properties as the parental bacteria, purification of the spores may not even be necessary.

Pharmaceutical and veterinary compositions for oral administration, comprising spores of transformed *Clostridia* and pharmaceutically acceptable carriers, diluents and excepients are further provided by the invention. They have the ability to colonise the intestinal tract of humans and animals with live *Clostridial* bacteria producing and presenting or secreting the heterologous polypeptide coded for by the modified gene cassette provided by the invention or by any other construct. The pharmaceutical and veterinary compositions may comprise tablets, capsules, powder for reconstitution or any other form suitable for oral administration to humans or animals. Examples of pharmaceutically acceptable carriers and diluents are lactose and carboxymethyl cellulose. A convenient way of oral administration of these therapeutic agents is to provide them as lyophilized, or just dried, powders; shortly before administration they are suspended in for example water, physiological saline or fruit juice. The dose is as indicated above for *Clostridial* bacteria.

In a further aspect of the invention we provide a method of treatment for the human or animal body, which comprises of administering a therapeutic agent comprising *Clostridial* spores capable of expressing a heterologous polypeptide in a human or animal body.

As in the treatment with *Clostridial* bacteria described above it will be possible to treat with mixtures of transformed *Clostridial* spores to obtain in the body a mixture of different heterologous polypeptides, which may act synergistically or additively.

It will be appreciated that the therapy may be either prophylactic or therapeutic.

The method may be applied to any convenient mammal such as a human or animal. Convenient animals include domestic animals such as dogs and cats, also cattle, pigs, chicken and horses.

In a further aspect of the present invention we provide a method for immunisation which method comprises administering to a mammalian body *Clostridial* spores capable of expressing a heterologous antigen after germination.

Examples of convenient *Clostridium* spores include the spores of *C. difficile* and *C. perfringens,* which normally colonise the large intestine of man. For animals other *Clostridia* such as *C. tetani* may also be useful. Thus, the intensity and duration of antigen exposure in the gut (clostridial colonization) in a particular host can be varied by not only exploiting and manipulating e.g. adherence of *C. difficile* (see above), but also by selecting the appropriate *Clostridium* species with regard to the intended host mammal.

In a further aspect of the invention we provide *Clostridial* spores transformed with a gene expression cassette of the invention.

It will be appreciated that the methods and materials of the invention may also be used for other applications such as the display of antibodies and peptide libraries. They may also be used for screening proteins and antigens and also to provide a support for immobilising an enzyme, peptide and/or antigen.

Sequence numbers of proteins encoded by the ORFs of the invention

Protein encoded by ORF	**SEQ ID NO:**
ORF 1	SEQ ID NO: 22
ORF 3	SEQ ID NO: 23
ORF 5	SEQ ID NO: 24
ORF 6	SEQ ID NO: 25
ORF 7	SEQ ID NO: 26
ORF 9	SEQ ID NO: 27
ORF 11	SEQ ID NO: 28

Protein encoded by ORF	SEQ ID NO:
ORF D	SEQ ID NO: 29
ORF E	SEQ ID NO: 30
ORF G	SEQ ID NO: 31
ORF H	SEQ ID NO: 32
ORF I	SEQ ID NO: 33

The present invention is particularly directed to a gene expression cassette comprising a secretory leader sequence encoding a signal peptide from *Clostridium difficile* having an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6 , SEQ ID NO: 7 and signal peptides of analogous exported clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked to a DNA sequence encoding a heterologous polypeptide. The signal peptides of the analogous clostridial N-acetylmuramoyl-L-alanine amidase-like proteins may also be selected from *Clostridium difficile* signal peptides having an amino acid sequence of any one of SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12.

The gene expression cassette may further include a promoter of prokaryotic origin, e.g. selected from clostridial promoters comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 13 – 21, or from the promoters of ORFs 1-11 or D-I mentioned above.

The gene expression cassette according to the invention may further include a DNA sequence encoding at least a cell wall binding portion of a protein of prokaryotic origin functioning in clostridia such that a fusion polypeptide may be presented on the outer surface of a host cell harbouring the cassette.

The gene expression cassette according to the invention may in particular include a DNA sequence encoding at least a functional cell wall binding portion of an S-layer protein of *C. difficile* selected from any one of the polypeptides having an amino acid sequence selected from SEQ ID NO: 22 - 33 such that a fusion polypeptide may also be presented on the outer surface of a host cell harbouring the cassette. The DNA encoding the cell wall binding portions of SEQ ID NO: 22-33 may be omitted such that the fusion peptide is secreted into the surrounding milieu by the host cell harbouring the cassette.

Further, the gene expression cassette according to the invention may be such that the DNA sequence encoding the heterologous peptide is inserted at a point downstream the first (signal) proteolytic cleavage sites in the gene encoding a polypeptide having an amino acid sequence selected from SEQ ID NO: 22 – 33, optionally including or excluding its second cleavage site.

In addition, the gene expression cassette according to the invention may further comprise at least a functional part of a secretory (*secA*) gene recognizing the signal peptide, to allow translocation of a heterologous polypeptide and/or fusion polypeptide across the cytoplasmic membrane of a host cell harbouring the expression cassette. For example, the secretory gene may be from *C. difficile* and encode a polypeptide having the amino acid sequence SEQ ID NO: 34.

In a preferred embodiment of the invention the gene expression cassette is the one that is shown in Figure 3.

The invention is also directed to a vector comprising a gene expression cassette according to the invention, such as a plasmid.

The invention is further directed to host organism transformed with a vector according to the invention for expression of the heterologous polypeptide and/or fusion polypeptide.

In an embodiment of the invention the host organism is a *Clostridium* host organism transformed with a vector according to the invention for expression of the heterologous polypeptide and/or fusion polypeptide.

In a preferred embodiment, the host organism is *C. difficile* or *C. perfringens*.

Further, the invention is directed to a pharmaceutical or veterinary composition or formulation which comprises *Clostridial* cells transformed with a vector according to the invention, with the ability to present on the cell surface and/or to secrete an expressed heterologous polypeptide or fusion polypeptide, together with a pharmaceutically or veterinary acceptable carrier or diluent. Preferably, the composition or formulation is suitable for oral or intranasal administration. The composition or formulation according to the invention may further comprise, as adjuvants, non-toxic motifs of the *C. difficile* toxins A and/or B that enable the heterologous polypeptide and/or fusion polypeptide to enter the colonic mucosal cells of a mammal by receptor-mediated endocytosis, and/or a portion of toxin B responsible for its intracellular and intercellular spread. The composition or formulation according to the invention may alternatively additionally comprise a further fused or separate carrier peptide or adjuvant, in addition to the expressed heterologous polypeptide

and/or fusion polypeptide, to elicit a stronger or differently directed immune response than that against the expressed heterologous polypeptide acting alone.

The invention is, in another aspect, directed to a vaccine which comprises a *Clostridial* bacterium transformed with a vector according to the invention and capable of presenting on the surface of the bacterium and/or secreting an antigen in a human or animal body, and optionally an adjuvant described in conjunction with a composition or formulation of the invention. The vaccine may comprise a mixture of at least two differently engineered *Clostridia* strains, each capable of presenting on the surface of the bacteria and/or secreting a different heterologous polypeptide and/or fusion polypeptide. Further, the vaccine may comprise spores of *Clostridia* cells or bacteria transformed with a vector according to the invention and capable of germinating into cells which are able to grow, express, and present or secrete a heterologous polypeptide and/or fusion polypeptide, and optionally also an adjuvant described in conjunction with a composition or formulation of the invention, in a mammalian body. The vaccine may comprise a mixture of spores from at least two differently engineered *Clostridia* strains. Each of these strains is capable of presenting on the surface of the bacterium and/or secreting a different heterologous polypeptide and/or fusion polypeptide. The spores are preferably from *C. difficile* or *C. perfringens.*

The invention is in yet another aspect directed to a medicament which comprises a *Clostridial* bacterium transformed with a vector according to the invention and capable of presenting on the surface of the bacterium and/or secreting a heterologous polypeptide and/or fusion polypeptide in a human or animal body, and optionally an adjuvant described in conjunction with a composition or formulation of the invention. The medicament may comprise a mixture of at least two differently engineered *Clostridia* strains, each capable of presenting on the surface of the bacteria and/or secreting a different heterologous polypeptide and/or fusion polypeptide. Further, the medicament may comprise spores of *Clostridia* cells or bacteria transformed with a vector according to the invention and capable of germinating into cells which are able to grow, express, and present or secrete a heterologous polypeptide and/or fusion polypeptide , and optionally an adjuvant described in conjunction with a composition or formulation of the invention, in a mammalian body. The medicament may comprise a mixture of spores from at least two differently engineered *Clostridia* strains. Each of these strains is capable of presenting on the surface of the bacterium and/or secreting a different heterologous polypeptide and/or fusion polypeptide. The spores are preferably from *C. difficile* or *C. perfringens.*

The invention is in still another aspect directed to a method for vaccination of a mammal, which comprises administering a therapeutically or prophylactically effective dose of a vaccine according to the invention to the mammal. Spores used in the vaccine are preferably from *C. difficile* or *C. perfringens.*

The invention is also directed to a method for prophylactic or therapeutic treatment of a mammal, which comprises administering a therapeutically or prophylactically effective dose of a medicament according to the invention to the mammal. Spores used in the medicament are preferably from *C. difficile* or *C. perfringens.*

The invention is additionally directed to a *C. difficile*-associated diarrhea (CDAD) vaccine comprising spores according to the invention and capable of expressing, after germination,

(i) relevant parts of the *C. difficile* toxins, alone or together with a

(ii) suitable adjuvant to provide an IgA response to the toxin antigenic epitopes and

(iii) S-layer protein antigenic variants (serotype antigens) or fimbrial antigens to obtain, after administration to a mammal, a polyvalent anti-S-layer (or anti-fimbrial) IgA response preventing *C. difficile* colonization of the mammal.

The invention will now be illustrated but not limited by reference to the following Figures, Specific Descriptions, Tables, and Examples wherein:

Figure 1A shows our first simple layout of the *C. difficile* strain 630 genomic segment encoding the S-layer genes. ORF2 represents *secA* and ORFs 1,3, 7-9 and 11 S-layer protein genes. For explanations of the these and other ORFs, see Table 1 and Example 9.

Figure 1B represents the result of comparisons between three of the S-layer ORFs with published sequences of other genes. The "amidase enhanced precursor" sequence is equivalent to the N-acetyl muramoyl L-alanine amidase motif mentioned in the text.

Figure 2. Defining the upstream region of ORF 1-12. The figure illustrates additional information and genetic organisation of the *C. difficile* S-layer genes (cf. Figure 1), found after searches in the revised *C. difficile* database at the Sanger Centre. The genes upstream of ORF 1 to 12 are denoted A to I (see also Table 1). The numbers +1, +2 and +3 indicate the reading frame of the ORFs relative to the start point of the contig. ORFs D, E, G, H and I had the amidase motif typical of genes encoding the *C. difficile* S-layer proteins.

Figure 3 shows an example of a preferred gene expression cassette here taken from *C. difficile* strain 630 and containing a strong promoter, the secretory leader peptide from ORF1, the signal peptide cleavage site the area of insertion of foreign DNA encoding the

heterologous peptide, the second (optional) peptide cleavage site in the N-acetyl muramoyl L-alanine amidase motif, and the *secA* gene (ORF2).

Figure 4 shows a preferred strategy for introducing a recombinant gene cassette of the invention back into *C. difficile via B. subtilis* .

Figure 5. Further details of a particular *C.difficile S - layer* gene cassette. This is a 4960 bp cassette taken from strain 630 encoding an S-layer protein of 2160 bp in its original form (ORF1). The 210 bp region (pr, promoter) upstream of ORF1 includes gene control elements for the S-layer protein included in the cassette. Also shown are an intervening 244 bp region and the 2346 bp *sec*A sequence.

Figure 6. Strategy for the engineering of ORF1 (Figure 5) to express a recombinant protein (for example as outlined in Example 2 and 2A). The 613 bp variable region (vr) is replaced by a foreign DNA. For example, three fragments encoding the Hepatitis B virus surface antigen (HBsAg) were selected: (i) the full length HBsAg that includes the pre S1, pre S2 and the S gene (1207 bp); (ii) the S gene (740 bp); and (iii) the subtype from the S gene (minimum antigenic epitope, 421 bp).

Figure 7. Cloning strategy for the construction of ORF1 - *secA* (with and without the native promoter) together with the different lengths of the HBsAg antigenic loop (full length, S gene and Sub type - see legend to Fig. 6 above) using a PCR based method and cloning into the TA vector in *E. coli.* The primers indicated were used also for PCRs to help checking the correctness of the constructs. The expected and obtained constructs were 5564, 5097 and 4778 bp respectively.

Specific descriptions

A. Characterisation of the genomic segment responsible for S-layer protein expression in *C. difficile* (outlined in Fig. 1 and Fig. 2, see also Table 1 and Example 9):

1. The main surface layer proteins expressed by *C. difficile* strain 630 has been found to encoded by a single open reading frame (ORF1) encoding a 72 kDa protein. The gene product of ORF1 is postranslationally cleaved at two sites yielding three different peptides; the leader peptide and the final S-layer proteins of apparent molecular weights of 36 kDa and 45 kDa.

2. The C-terminus of ORF1 shows similarity to N-acetyl muramoyl L-alanine amidase, and the N-terminus shows weak similarity to surface layer proteins from *L. helveticus* (Fig. 1B).
3. The gene immediately downstream of ORF1 (ORF2) encodes the SecA protein responsible for secretion of proteins with signal peptides.
4. The genes downstream of *secA* encodes proteins with similarities to ORF1 (ORF3, 5-7, 9 and 11, see Fig. 1).
5. ORF1 is efficiently expressed and its product is efficiently exported in strain 630, whereas e.g. ORF3 is expressed more than 100-fold less in strain VPI 10643, indicating a strong termination between ORF1 and 3 (as judged by identification of exported proteins by two-dimensional gel electrophoresis).
6. The upstream region including the putative promoter has not been characterized functionally, but the very high expression of ORF1 in various growth conditions indicates the action of a strong, constitutive promoter. It is also active in *E.coli* (see Example 2).
7. In the revised *C. difficile* sequence database from Sanger Centre, the upstream region of ORF 1 was included and revealed 9 new ORFs (A-I) of which 5 (D, E, G, H and I) had the N-acetylmuramoyl L-alanine amidase motif typical of the *C. difficile* S-layer protein ORFs (Fig. 2 and Table 1). The putative promoter region for ORF1 is thus situated between ORFI and ORF (See Example 1A and Table 2).

The S-layer proteins from strain VPI 10463 have similar molecular weights but different pI as compared to those of strain 630, and the N-terminal sequences of the two S-layer proteins from VPI 10463 showed no similaritiy with those of strain 630. Studies of strains from different serogroups showed that the S- layer proteins vary in pI and molecular weight. The downstream region of the gene segment may in part be more conserved, since the N-terminal sequence from another extracellular protein from strain VPI 10643 was identical to ORF3 of strain 630. Our results indicate that ORF1 is located at part of the chromosome that is capable of expressing and exporting various S-layer proteins depending on the strain.

B. Designing a preferred engineered gene cassette based on the *C. difficile* S-layer gene segment data (outlined in Fig. 3).

1. Any strong prokaryotic promoter functional in *Clostridia* can be used to express the heterologous peptide, e.g. the promoter of ORF1 or any of the promoters of genes encoding

other highly expressed proteins in *C. difficile* such as certain electron transfer proteins (our unpublished data and Fix A and Fix B in Example 9) or ribosomal proteins.

2. A secretory leader peptide, preferably the leader peptide from ORF1, is fused with the heterologous peptide, to ensure its translocation across the cell membrane

3. Depending on the desired fate of e.g. the antigen (secreted, surface presented or both), the heterologous peptide is optionally fused to the amidase part of the S-layer protein optionally including the part involved in the proteolytic cleavage event (Figure 3). Thus, for maximum release of the peptide the secretory leader of e.g. ORF1 may be sufficient. On the other hand, maximum cell-wall binding may require fusion to the amidase portion but omitting the proteolytic cleavage sequence in the middle of the gene (Figure 3). If both free and bound heterologous peptide is desired one recombinant cassette of each type present in the same *Clostridium* strain or a mix of two different strains, each harbouring one the recombinant cassettes, can be used. The peptide cleavage site may be exploited if for instance the antigen and an adjuvant are produced in a fused form, to obtain equal amounts of the two, but are desired as separate peptides on the outside of the producer bacterium. To what extent parts of the N-terminal (variable) portion of e.g. ORF1 can be used to optimize the localization of the heterologous peptide requires further experimentation.

4. The *secA* gene is usually included in the construct to ensure efficient translocation of the polypeptide across the cell membrane.

5. The gene construct is made in plasmids suitable for transformation of both *E.coli* and *C. perfringens* (e.g. pJIR750 or 751) or in plasmids suitable for conjugation into *C. difficile* via *B. subtilis* (e.g. pCI195 or pSMB47, Figure 4).

Examples

Example 1

Defining a minimal gene segment responsible for production and export of the heterologous protein (an engineered gene cassette, see also Figure 3).

A. Defining the promoter start (the 5' end).

The promoter region may be further characterised in different *C.difficile* strains, for example by the following steps:

1. Defining the DNA upstream of ORF1 by circular PCR. One or several restriction enzymes are used to cut the DNA outside ORF1. Useful enzymes are NdeI, TaqI, PstI, and BsmAI.

A Southern blot is performed using a probe against ORF1 to confirm size of the fragment (optimal size is around 1-3 kb). The cleaved genomic DNA is then ligated into circles. PCR is then performed using primers (directed outwards of each other) directed against ORF1. The PCR product is then cloned and sequenced. Optionally, PCR is performed with one primer directed against ORF1 (lower primer) and different arbitrarily designed primers. The products are then cloned and sequenced.

2. Identification of the transcription start. Primer extension with primer located at the 5 prime end of the genomic segment is used.
3. By computerised search for homologies to known promoters. The new sequence data for strain 630 enabled a search for putative promoters upstream of ORF1 and the result is shown in Table 2. Due to the AT rich genome, several putative promoters were found. The actual promoter start has to be experimentally determined as described in point 2 above.

B. Defining the termination point (the 3' end).

Primers directed at different parts of ORF1, 2 and 3 to determine transcript abundance using Northern blots. Termination loops in the RNA may be identified by computer analysis.

C. Conservation of the S-layer locus in other *C. difficile* strains: characterization of part of the S-layer genomic segment (ORF1 and proximal parts of its upstream region and of part of ORF2) in different *C. difficile* serogroups in order to check for conservation of this region of the genome.

Primer pairs that are directed against the identified upstream-transcription start region and the proximal part of ORF2 were designed. PCR was performed on different strains belonging to all serogroups to confirm the generality of the expression center of the S-layer locus.

(i) PCR performed with chromosomal DNA as template from the different serogroups of C. *difficile.*

The primers used:

CONS1: 5 prime- TAT AAT GTT GGG AGG AAT TTA AGA – 3 prime, total length 24 nt
(5 prime end starts at 8th nt upstream of ORF1, ends at 32nd nt)

CONS2: 5 prime- CAA ATC CAA ATT CAC TAT TTG TAC – 3 prime, total length 24 nt
(5 prime end starts at 2983rd nt downstream of ORF1, ends at 2959th nt)

Total size of expected PCR pdt (from strain 630 sequence): 2975 bp (includes ORF1 and the proximal part of ORF2).

The enzyme/system used: Expand™ Long Template PCR System from Boehringer Mannheim.

Reaction conditions (as specified by the manufacturer): In a total reaction volume of 50ml, 350mM dNTPs, 300nM primers, 50ng chromosomal DNA template, 1x supplied PCR buffer with 1.75mM $MgCl_2$, and 2.5U of a mix of Taq and Pwo DNA polymerase. 10ml of the reaction mix was run on a 0.8% Agarose-TBE (Tris-Borate- EDTA buffer) to check for product.

PCR cycle conditions: Initial denaturation - 92°C for 2 mins

For 30 cycles - denaturation - 92°C for 10 secs

annealing - 40°C for 30 secs

elongation - 68°C for 2 mins

Final elongation - 68°C for 2 mins

PCR results:

Serogroup tested	Product size	No. of Expts
A	none	3
B	around 2900bp	2
C	around 2900bp	2
D	around 2900bp(lesser amt)	2
F	around 2900bp	2
G	around 2900bp	2
H	pdt > 2900bp (lesser amt)	3
I	around 2900bp	2
K	around 2900bp	2
X	around 2900bp	2
A2	around 2900bp (lesser amt)	3

Serogroup tested	Product size	No. of Expts
A3	around 2900bp (lesser amt)	2
A4	around 2900bp (lesser amt)	2
A5	none	2
A6	around 2900bp (lesser amt)	2
A8	around 2900bp	2
A9	around 2900bp (lesser amt)	2
A10	around 2900bp	2
S1	around 2900bp	2
S3	around 2900bp	2
S4	none	2
VPI 10643 (G)	around 2900bp	4, last PCR was weak
630 (X)	around 2900bp (expected size)	5, last PCR was -ve

Serogroups A, A5 and S4 did not give any PCR product with several attempts.
The PCR reactions were very sensitive to template condition, which had to be prepared fresh.

(ii) Restriction Enzyme Analysis of the PCR products.

12ul PCR reaction mix (reactions done above) from most of the serogroups (except A, A5-6, A9, and S4) were further subjected to RsaI (Boehringer Mannheim) and Sau3A1 (Amersham Pharmacia Biotech) digestions. Total digestion reaction mixes were run on a 0.8% Agarose-TBE (Tris-Borate- EDTA buffer) to obtain respective digestion patterns.

Each serogroup tested appears to give rise to unique restriction pattern though the apparent size of the PCR product appears to be similar except for gp H.

Conclusion:

It appears that the promoter-ORF1-ORF2 organization and the size of the genetic segment between the putative promoter and ORF2 (*secA* gene), i.e. the ORF 1 equivalent, is generally conserved in *C. difficile*. There may be sequence variation between ORF 1 eqivalents from strains of different serotype and also between two strains of the same serotype, presumably reflecting the variable (non-amidase) portion of these genes.

Example 2

Cloning of the *C. difficile* promoter-ORF1-ORF2 "cassette" from strain 630 and construction of a recombinant cassette in *E. coli* encoding heterologous proteins to be transferred to *C. perfringens* and *C. difficile* and used for immunisation .

1. Primer pairs are designed that include the promoter region and part of ORF1 including the leader peptide sequence (Figure 5). PCR is performed followed by cloning of the product into *E. coli-C. perfringens* shuttle vectors pJIR750 or pJIR751 (Plasmid 229: 233-235, 1993) in frame of a reporter gene such as β-lactamase or at least a part of the hepatitis B virus (HBV) antigen. A convenient source of HBV antigen is the SMI strain no. 8423/87 having the genotype A and subtype adw2 (cf. Magnius et al, J.Gen. Virology, 1993, 74, 1341-1348). The plasmid is isolated from *E. coli*, purified and used to transform *C. perfringens*, and the engineered strain is isolated for further use. The *secA* gene is optionally included in the construction to optimise secretion.
2. Check for expression and secretion of the reporter gene using an HBV antigen based assay or a β-lactamase assay of the transformed *E.coli* and *C. perfringens* strain. Gnotobiotic mice and rats are fed with spores of the engineered *Clostridium* strain to obtain colonisation. Expression of antigen is checked by the HBV antigen based assay or β-lactamase assay in feces and immune response by antibody response in feces and in serum.

A. Cloning of ORF1 - ORF2 (*secA*) in order to construct a fusion with a foreign antigen

PCR was performed with chromosomal DNA from Strain 630 as template. The primers used:

AMP1: 5' - GGA ATT CCA TGA ATA AGA AAA ATA TAG CA - 3',
total length 29 nt (5'end starts at the first codon of ORF1, ends at 7th codon)

AMP2: 5' - CGG GAT CCC GTT TTT AGT TAA ATT TAT ATA AG - 3',
total length 32 nt (5'end starts at starts at the stop codon for *secA*)

Analogous PCRs but with the first primer in the upstream region in order to include a putative native promoter were also performed.

Total size of expected PCR product (from strain 630 sequence) : 4770 bp (4960 bp including the promoter) (Figure 5).

The enzyme/system used: Expand™ Long Template PCR System from Boehringer Mannheim. Reaction conditions (as specified by the manufacturer): In a total reaction volume of 50ul, 350mM dNTPs, 300nM primers, 150ng chromosomal DNA template from strain 630, I x supplied PCR buffer with 1.75mM $MgCl_2$, and 2.5U of a mix of Taq and Pwo DNA polymerase. 10ul of the reaction mix was run on a 0.8% Agarose-TBE (Tris-Borate- EDTA buffer) to check for product.

PCR cycle conditions: Initial denaturation - 92°C for 2 mins

For 30 cycles - denaturation - 92°C for 10 secs

annealing - 40°C for 30 secs

elongation - 68°C for 4 mins

final elongation - 68°C for 5 mins

Results:

The expected PCR products were obtained and cloned into pGEMT vector (Promega). The plasmid containing the insert will be subjected to partial digestion with PvuII enzyme (sites at position 282 and 895 of the insert) to eliminate the 613 bp internal fragment from ORF1, where the foreign antigen is planned to be inserted (Figure 6). The digestion time had to be standardised. The foreign antigen used was the hepatitis B virus (HBV)surface antigen (HbsAg).

Three HBsAg regions were used (Figure 6): (a) Total (413 aa, 3-1243 bp; 1240 bp)

(b) Central (259 aa, 465-1243 bp; 778 bp)

(c) C-terminal (140 aa, 822-1243 bp; 421 bp)

References for HBV antigen sequences include Prange et al, 1995, 14 (2), 247-256 and Chen et al, 1996, 93, 1997-2001.

Alternative antigens that may be used include relevant epitopes of the rota virus and hepatitis A virus.

B. Cloning of the *C. difficile* ORF1-*secA* cassette containing the three HBsAg DNAs in *E. coli*.

Result:

The cloning strategy (Figure 7) using PCR and the TA vector was successful according to DNA analyses including agarose gel electrophoresis and PCR.

C. ELISA to check for expression in *E. coli* of the three ORF1-HBsAg-*secA* DNA constructs in the TA vector.

A commercially available ELISA (Abbott) was performed on sonicated samples of three overnight cultures of *E. coli* each containing one of the three different HBsAg DNAs inserted into ORF1-*secA* (with the native *C.difficile* ORF1 promoter) and cloned into the TA vector.

Results:

Cut off value = 1.0, higher values are regarded as positive.

Sample (plasmid)	*Culture OD**	*Sonicate-supernatant (periplasm +cytoplasm)*	*Pellet following sonication (cell membrane)*	*Spent medium*
1.3	1.1#	0.92	Test Failed	0.61
1.3a	1.9	4.76	3.70	0.69
2.8	1.6#	0.99	1.33	0.74
2.8a	1.8	1.14	4.33	0.69
3.7	1.4#	18.56	34.58	0.70
3.7a	2.0	5.10	5.11	0.69
TA vector	2.0#	0.70	1.13	0.69
TA vector a	2.0	0.74	0.91	0.70
PBS(buffer)	NT	0.77	NT	NT

1.3 : Full length HBsAg(pre S1, pre S2 and S gene),

2.8 : S gene,

3.7 : Partial S gene(the minimum HBsAg epitope)

a : Duplicates

: Samples stored overnight in cold before analysis.

NT : Not tested

***** : Culture OD at the time of harvest. Cells from 1ml culture were pelleted and resuspended in 0.5ml of PBS before sonication.

The above experiment was repeated with cells from 5ml cultures resuspended into1ml PBS to see if increased protein concentrations would give higher titer values. Here a control *E. coli*

culture with the TA vector carrying the ORF1-*secA* cassette but without any HBsAg insert was also included.

Sample	*Culture OD*	*Sonicate-supernatant (cytoplasm)*	*Pellet following sonication (cell membrane)*
1.3	2.4	8.17	3.08(1/10 dil)
2.8	1.8	5.07	2.02(1/10 dil)
3.7	2.4	42.0	1.15(1/100 dil)
orf1secA (no HBsAg insert)	2.4	Test Failed	0.92

Concusion:

All three constructs containing HBV DNA expressed both soluble and and particulate HBsAg. Thus, the cloning proved to be successful and the native *C. difficile* ORF1 promoter was to some extent functional also in *E. coli.*

Example 3

Cloning of the ORF1-ORF2 (*secA*) "cassette" into *C. difficile* and construction of recombinant protein useful for immunisation studies.

1. Cloning of a gene expression cassette according to examples 1 and 2 in plasmids pCI195 or pSMB47 followed by transfer to a non-toxigenic strain of *C. difficile* by mating via *B. subtilis* is performed according to standard methods reported (J. Gen. Microbiol 136: 1343-1349, 1990; Plasmid 31: 320-323, 1994, see also Figure 4). The engineered strain is isolated for further use.
2. Check for expression, secretion and antibody response *in vitro* and *in vivo* (see Example 2).

A. Cloning of the ORF1-HBsAg-*secA* constructs into the shuttle plasmid vector pJIR750

Results:

Ligation mixtures containing the desired recombinant plasmids were obtained an judged by agarose gel electrophoresis and PCR. However, upon transformation into *E. coli* the plasmid constructs were fragmented. This indicated that a plasmid replication machinery better at handling large plasmids in *E. coli* than that of pJIR750 (colE based) needs to be used. Also,

attempts to transform *C. perfringens* with our recombinant pJIR750 plasmids are being performed.

Example 4

Production of transformed *C. perfringens* expressing and presenting foreign antigen for vaccination.

A. Live bacteria.

C. perfringens, transformed with a gene cassette coding for a foreign antigen fused to ORF1 and obtained as in Example 2 is cultivated under anaerobic conditions in a fermenter until a cell density of at least 10^7 bacteria per ml is obtained. The broth is cooled to 11°C, the bacteria recovered by centrifugation and the supernatant discarded. The pellet is twice washed with cold 0.1 M phosphate buffer, pH 7 and centrifuged. The final pellet is resuspended in the phosphate buffer to a concentration of about 10^9 organisms per ml. One ml portions of the suspension are dispensed into glass ampoules and freeze-dried to remove the water. The final product is obtained by sealing of the ampoules *in vacuo*.

B. Bacterial envelopes.

Transformed *C. perfringens* bacteria are produced as in A above. The final pellet is suspended in 50 mM Tris-HCl, pH 7.2, and sonicated for 1-10 min (40 watt, Bransic Sonic Power co. Sonicator). Triton X-100 is added to a final concentration of 2% and the mixture incubated under stirring at 11°C for 30 min.. The cells are collected by centrifugation and washed three times with cold distilled water. The pellet is resuspended in 5 mM MgCl, containing DNase (11mg/ml) and RNase (11mg/ml) and incubated for 15 min at 11°C. The resulting envelopes are recovered by centrifugation, washed three times with cold distilled water, resuspended in cold distilled water and freeze-dried to give the envelopes as a powder suitable for formulation in capsules or tablets, for suspension in e.g. physiological saline for oral administration.

Example 5

Production of Clostridial spores, germination of spores *in vitro* and *in vivo*, and colonization and immune response to *C. difficile* in animals given spores orally

Clostridial strain producing the heterologous peptide is allowed to grow anaerobically in Peptone-Yeast extract-glucose or another medium optimal for sporulation for 48-72h to ensure maximum conversion of the vegetative bacteria into spores during the stationary phase. The remaining vegetative bacteria are killed by heat or ethanol treatment, eliminated by the bacteriolytic enzymes lysozyme or lysostaphin and the remaining spores are purified by centrifugation.

Results:

Spores from *C. difficile* strain 630 were readily observed on 3-day old blood agar plates as well as in PYG medium, whereas spores from *C. perfringens* NCTC 8798 were found only during growth in Duncan Strong (DS) medium, after preinocubation in Fluid Thioglycolate (FTG) medium.

Prepared spores (washed in ethanol and resuspended in sterile PBS) were checked for germination ability on plates (TCCFA) as well as in gnotobiotic rats. Feces from 1 week old rats were positive for bacterial growth in feces 1-2 days after receiving *C. difficile* or *C. perfringens* spores orally, confirming the ability of the spores to germinate also *in vivo* and lead to colonization of the animal gut.

Antisera from 5 rats colonised for one week by *C. difficile* were pooled and used for Western blotting of *C. difficile* protein extracts. Western blotting revealed immunological reactions to bands corresponding to the *C. difficile* S-layer proteins confirming that antibodies were produced against these *C. difficile* antigens upon feeding with spores, spore germination and colonization of the animals.

Example 6

Production of transformed *C. difficile* spores for use in oral immunisation.

A. Capsules.

The *C. difficile* spores obtained according to Example 5 are mixed together with Mg stearate (1%) and lactose (30%), granulated in ethanol and compressed to tablets, containing 10^6 spores, which are filled into gelatine capsules.

B. Powder for suspension.

The *C. difficile* spores obtained according to Example 5 are mixed together with Mg stearate (1%) and carboxymethyl cellulose (25%), granulated in ethanol. The granulate is dried and dispensed into vials to give an amount of about 10^6 spores in each vial. For oral

administration the content of the vial is suspended in water or for example orange juice immediately before intake.

Example 7

Use of S-layer genes for epidemiological typing.

Present methods to detect and follow the spread of certain *C. difficile* strains in the environment and between infected patients (i.e. "fingerprinting") include e.g. serotyping and PCR ribotyping. PCR ribotyping is a PCR based approach to amplify the region between the 16S and 23S genes of *C. difficile*, and which has been shown to resolve and detect over 100 different patterns or strains. Different serotypes are likely to represent differences of the surface-exposed proteins, i.e. variations of S-layer proteins among strains. Our results with PCR amplification and following cleavage with restriction enzymes indicate that this region is present in almost all serogroups and that the cleavage pattern also varies among these groups (see Example 1C, (ii)). Thus, a molecular method including PCR combined with restriction enzyme cleavage or direct sequencing of the variable part of the ORF1 or another part of this segment may be a method which is faster and more reliable than serotyping and in particular also more reliable than PCR ribotyping for fingerprinting.

Example 8

Vaccination against CDAD

Immunity to CDAD after an episode of the infection is regarded to be short (months). This may be due to that anti-toxin antibodies are mainly of the serum IgG classes and not the secreted IgA class made to protect the gut mucosal surface, because the toxins are internalized by the gut mucosal cells (see above) and not by the M-cells specialized in furthering an IgA response. A further problem may be that immunity to the 20 *C. difficile* S-layer serotypes is required for prevention of colonization and thus the best protection against infection. For these reasons, it is likely that injectable vaccines against CDAD based on the toxins and under development may turn out to offer poor protection.

As an alternative, we provide a polyvalent live oral vaccine containing (i) the most prevalent toxin-producing serotypes (S-layer variants), here attenuated by knock-out of their toxin genes, and (ii) carrying a recombinant ORF1-*secA* cassette encoding relevant parts of the toxin genes and (iii) an adjuvant peptide ensuring uptake of the immunogenic toxin epitopes by e.g. M-cells in order to obtain an IgA anti-toxin response.

Example 9

Analysis of extracellular and membrane fraction proteins in *Clostridium difficile* by two-dimensional PAGE, N-terminal sequencing and data base searches – identification of genes encoding the S-layer proteins

Identification of extracellular proteins

Analysis of its extracellular protein pattern by 2-D PAGE showed that two proteins of 50 kDa and 36 kDa were very abundant. Subsequent analysis of membrane preparations from *C. difficile* VPI 10643 corroborated the almost exclusive fractionation of the 50 kDa and 36 kDa proteins into the membrane fraction, when compared with the soluble fraction. Also in strain 630 two proteins with similar molecular weights but with different pI were abundant in the extracellular as well as in the membrane fraction. Thus, the 50 kDa and 36 kDa proteins were likely to constitute the *C. difficile* surface(S)-layer proteins, which are known to be partially shed from the bacterial surface into the culture supernatant (Luckevich and Beveridge, 1989; Tsukagoshi *et al.*, 1984). The N-terminal sequence of spot no. 1 from VPI 10463 did not show any homology to other proteins in the *C. difficile* strain 630 genome database (Table 3). The N-terminal part of spot no. 2 showed similarity to an open reading frame encoding a 72 kDa protein in the *C. difficile* genome database (Table 3; ORF1, see also Figure 1). However, only nine out of 15 amino acids matched close to the N-terminus of ORF1. Strikingly, the N-terminal sequences of the corresponding proteins from strain 630 were different from those of VPI 10463 and both matched to ORF1 but at two different positions (spot No.10 and 11 in Table 1).

Several proteins were specifically found in PY cultures, i.e. during high toxin production (Table 3, spot no. 3, 4, 5, and 6). The N-terminal sequence of spot no. 3 matched with an ORF of 47.5 kDa in the *C. difficile* genome database. This ORF showed weak homology to a hypothetical protein in the *Plasmodium falciparum* genome database. The N-terminal sequence of spot no. 4 matched with an ORF of 39 kDa that showed homology to a phage-like element PBSX protein (XkdK) from *Bacillus subtilis*. The N-terminal sequences of spot no. 5 and 6 matched with two ORFs of 38 and 22 kDa, respectively, and these ORFs had the highest similarity to the FixB and FixA proteins from *Escherichia coli*. Spot no. 7, 8 and 9 were absent in PY but abundant in culture medium from PYG cultures. Spot no. 7 and 8 matched to an ORF located on the same contig as ORF1 (Table 3; ORF3). The N-terminus of

spot no. 9 matched to a central part of ORF1, and is likely to be a proteolytic fragment of a protein encoded by ORF1.

Analysis of the surface layer genes

The identification of the S-layer genes revealed a genomic segment including seven genes (ORF1, 3, 5-7, 9 11) with significant homology to N-acetyl muramoyl L-alanine amidase (CwlB/LytC) and modifier protein of major autolysin (LytB) from *Bacillus subtilis* (Fig.1, Table 4). In addition to the LytB/LytC similarity, the N-terminal part of ORF6 showed similarity to eukaryotic cysteine proteases, and the highly expressed ORF1 (above) showed weak similarity to S-layer proteins from *Lactobacillus* and *Streptococcus* spp. (Fig. 1). The N-terminus of ORF1 contained a typical signal peptide for export via the Sec-dependent secretion and the predicted cleavage site was identical to that found in the protein sequence (not shown). However, no typical protein cleavage site was identified within ORF1 that would allow processing of the 72 kDa protein further to give the finally sized S-layer proteins found (50 and 36 kDa). Strikingly, no significant match between the *C. difficile* S-layer ORFs and the S-layer homology motif (SLH domain) found in all presently known S-layer proteins was obtained (not shown). Most of the remaining genes in this genomic segment showed similarity to genes involved in secretion, polysaccharide and capsule synthesis (Fig. 1; Table 4). At least 2 other genomic sequence segments were found that contained genes similar to CwlB/LytC, indicating a complex variability (not shown).

Summary of results

The most dominant surface-exposed protein in many bacterial species is the S-protein. This protein crystallizes into a regular monolayer on the outside surface of the bacteria: the S-layer. The S-layers satisfy multiple roles for the cell and function as protective coats, as structures involved in cell adhesion and surface recognition, as molecular seives, as molecular and ion traps, as scaffolding for enzymes and as virulence factors (Sleytr and Beveridge, 1999; Sara and Sleytr, 2000). Though all S-layers share general features (all are made of relatively large proteins, self-assemble and are paracrystalline), comparative studies indicate that S-layers are non-conserved structures and are of limited taxonomical value (Kuen and Lubitz, 1996; Sleytr et al. 1999). Chemical analysis and genetic studies of a variety of S-layers have shown that they are composed of a single, homogenous protein or glycoprotein species with molecular weights ranging from 40 to 170 kDa. The S-layers of *Clostridium*

difficile (Takeoka et al., 1991) and *Bacillus anthracis* (Mesnage et al., 1998; Etienne-Toumelin et al., 1995) consist of two types of S-layer subunits which together form a defined lattice type but do not cross-react with polyclonal antibodies. Typically, S-layer proteins are often weakly acidic proteins (pIs between 4 to 6), containing 40-60% hydrophobic amino acids, and possess few or no sulfur-containing amino acids (Messner, 1996). S-protein production is directed by single or multiple promoters in front of the S-protein gene, yielding stable mRNAs. Most bacteria secrete S-proteins via the general secretory pathway (*sec*-pathway). Silent S-protein genes have been found in *Campylobacter fetus* and *Lactobacillus acidophilus*. These silent genes are placed in the expression site in a fraction of the bacterial population via inversion of a chromosomal segment (Boot and Pouwels,1996).

The S-layer has been detected in some *C.difficile* strains and preliminary characterization has been done from *C.difficile* C253 (Mauri et al., 1999). Another independent study purified and identified the S-layer subunits from *C.difficile* GAI 0714 (Takeoka et al., 1991). In both cases, the S-layer has been shown to be composed of two different protein subunits with apparent molecular weights of 36 kDa and 47 kDa (*C.difficile* C253) and 32 kDa and 45 kDa (*C.difficile* GAI 0714). The S-layer proteins from VPI 10463 and strain 630 was here found to be similar in size but with significant pI differences. The N-terminal sequences varied significantly especially for the larger protein. The N-terminal sequences as determined for these proteins also indicate that they are not identical. Those from strain 630 appear to be processed products from the same gene (ORF1, Table 3). The N-terminal sequences of spot 1 from VPI 10643 did not find any homologue in the strain 630 database corroborating earlier results that the higher molecular weight S-layer protein was sero-group specific (VPI 10643, group G and strain 630, group X) (Takeoka et al., 1991). Our current work indicated that this arrangement of two S-layer proteins was also true for all the different serotypes tested (data not shown). The spot 2 found a partial match with another ORF (ORF3) in the same contig as ORF1.

Other ORFs located in the same contig also had similarities with ORF1 and ORF3, whose C-terminal parts showed similarities to N-acetyl muramoyl L-alanine amidase (CwlB/LytC) and modifier protein of major autolysin (LytB) from *Bacillus subtilis* (Lazarevic et al., 1992), whereas the N-terminal part showed weak similarities to surface-layer proteins from *Lactobacillus helveticus* (Callegari et al., 1998) and *Streptococcus* spp. It is interesting to note that N-acetyl muramoyl L-alanine peptidoglycan amidase is the major autolysin of *B. subtilis* and has high affinity for cell walls, which is enhaced by the modifier protein, but

small amounts of cell free autolysin can be detected in cultures of *B. subtilis*. Thus, the amidase-like motif that appears to be typical of *C.difficile* S-layer proteins probably confers their anchorage to the cell wall peptidoglycan-teichoic acid.

Considering the highly competitive situation of closely related organisms in their natural habitats, it is obvious that the S-layer surface has to contribute to diversification rather than to conservation. With respect to this, the importance of S-layer variation leading to the expression of alternative S-layer genes under different stress factors such as those imposed by the immune system of a host in response to an S-layered pathogen or drastic changes in the growth and environmental conditions for nonpathogens is conceivable (Dworkin and Blaser, 1997; Sara et al., 1996). This could probably explain the variation in S-layer proteins even amongst the same species as in the case of *C. difficile.*

Identification of spot 4 as having similarity with XkdK, a protein encoded by the phage-like element PBSX from *Bacillus subtilis* (Krogh et al., 1996) and being located in a contig with other ORFs having similarity with other PBSX encoded proteins is very interesting. PBSX is a bacteriophage-like bacteriocin, or phibacin, of *B. subtilis* 168 (Okamoto et al., 1968). When *B. subtilis* 168 cells are exposed to treatments that induce the SOS response (such as UV light, mitomycin C), the cells lyse after incubation of 1h and release particles of PBSX (Seaman et al., 1962). The spot 4 is completely absent in PYG supernatants. Taken together this could indicate that toxin production (high in PY) and expression of this phage-like protein in *C. difficile* is a response to certain stress, environmental or otherwise, that decides whether it will resort to toxin expression, sporulation or both.

The N-terminal sequences of spots 7 (41 kDa) and 8 (38 kDa) are identical (Table 3) and correspond to the same ORF (ORF3, Fig 1), whose N-terminal part is similar to N-acetyl muramoyl L-alanine amidase (CwlB/LytC) from *B. subtilis*. However, the size of the proteins in the gel do not match the size expected from the ORF3 (encodes a 67.5 kDa protein). Both ORF1 and ORF3 have clear signal sequences at the beginning which is missing in the protein spots sequenced, thus indicating that these are indeed secreted and processed following translation. This could also possibly explain the difference in size and pI for the two different spots, and the discrepancy between expected and observed molecular weights on SDS-gels.

The spot 9 (24 kDa) has a N-terminal sequence (Table 1) which corresponds to an internal fragment of ORF1. The expected size of this fragment is around 21 kDa which corresponds closely with what is observed experimentally. Clearly there are post-translational processing events which could be enacted in the cell envelope or in the supernatant. It is

important however to note that spots 7-9 are present in PYG supernatants only, when the cells start sporulating.

The spots 9 and 10 are also processed products from ORF1 and are present in both PY supernatant and membrane fractions. However, these samples are obtained from strain 630. The results obtained thus far indicate that this operon (contig) (Fig 1) is present in both VPI 10643 and strain 630, but different ORFs are expressed by the two strains.

Experimental procedures

Strains and growth medium

The toxin-producing *C. difficile* strain VPI 10463 (CCUG 19126, Culture Collection, University of Göteborg, Sweden) was grown in either PY, PYG (purchased from the Karolinska Hospital, Stockholm, Sweden) or SDM medium. SDM is identical to MADM (Karasawa *et al.*, 1995; Yamakawa *et al.*, 1994; Yamakawa *et al.*, 1996), except that the concentrations of glycine and threonine were 100 mg/L and 200 mg/L, respectively, and that Ca-D-panthotenate, pyridoxine and biotin were used as the sole vitamin sources. PY(G) was prepared by adding cysteine (500 mg/L), boiling for 20 min while purging with an anaerobic gas mixture (10% CO_2, 10% H_2, 80% N_2) for 20 min, sterilised by filtration (Acrodisc, Gelman sciences) and aliquoted into tubes with serum vial-style necks (Bellco Glass) while flushing with anaerobic gas. The tubes were closed with butyl stoppers secured with aluminium crimp seals. SDM was prepared accordingly.

Growth conditions, sampling and optical density measurements

For each experiment, a tube containing 20 ml SDM was inoculated with 0.2 ml thawed bacterial suspension (stored at -70°C) using a syringe and a needle that was passed through the rubber septum of the tube. To avoid entry of O_2, the syringe was equilibrated with anaerobic gas before inoculation. The tube was put horizontally on a rotary shaker (50 rpm, 37 °C), and on the next day, the culture was serially diluted into PY or PYG. On day three, samples were collected from the diluted cultures and OD was measured at 600 nm using a Hitachi U-1100 spectrophotometer.

Sample preparation and membrane fractionation

Culture samples were centrifuged at 16000 x *g* for 3 min, whereafter the supernatants were removed, filtered, and stored at -20°C for later analysis. The pellet was frozen at -20°C

for 30 min or longer, thawed, dissolved in 1 ml sterile water and sonicated on ice for 3 x 30 s at 100 W (Labsonic 1510, B. Braun). Larger cell pellets, obtained from >1 ml culture, was sonicated for longer times. The cell extracts were centrifuged at 5000 x g for 5 min. The pellet was separated as the low speed pellet (LSP), and the supernatant was further centrifuged at 50000 x g for 20 min. The pellet was separated as the high speed pellet (HSP), and the supernatant (soluble fraction) was stored at -20°C. The LSP and the HSP were resuspended in 1x PBS (Phosphate buffered saline). Protein amount was measured using a kit (Biorad) and a BSA standard curve according to the manufacturerís instructions.

The culture supernatants were precipitated using trichloroacetic acid (TCA) to a final concentration of 10%. The pellets were washed with ice-cold Acetone, air dried and finally resuspended in 1x PBS to obtain the extracellular protein fraction. Protein estimation and analysis was carried out as described earlier.

Immunoprecipitation

Immunoprecipitation was performed in microtiter wells coated with antibodies against toxin A (PCG-4, r-Biopharm) or toxin B (xxx, r-Biopharm), Ten µg/ml antibody in 0.04 M Na_2CO_3, 0.06 M $NaHCO_3$, pH 9.6 was added to microtiter wells and incubated for 1 h at 37°C and washed four times with PBS containing 0.05% (v/v) Tween-20, pH 7.4. The wells were loaded with cell extract, culture supernatant medium or PBS (negative control), incubated 90 min at 25°C, and washed four times with PBS. After addition of 50 µl SDS sample buffer solution (below) and heating for 5 min at 95°C, the precipitated proteins were analysed by SDS-PAGE.

SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed using pre-cast polyacrylamide gels (ExcelGel 8-18% gradient gels, Pharmacia Biotech) and a Multiphor II horizontal slab gel apparatus (Pharmacia Biotech) according to the manuals provided by the manufacturer. The samples were mixed 1:1 with SDS sample buffer solution (0.05 M Tris, 1% (w/v) SDS, 10 mM DTT, 0.01% (w/v) bromophenol blue, pH 7), incubated 5 min at 95°C, loaded onto the gels and run at 15°C. Chemicals were obtained from Sigma, and molecular weight markers from Pharmacia Biotech. The gels were stained with silver (PlusOne, Pharmacia Biotech) using a Hoefer automatic gel stainer (Pharmacia Biotech), digitised by scanning (Scanjet 3c/T, Hewlett-Packard), and transferred to ClarisDraw (Claris Software) on a Macintosh computer.

Immunoblotting

Proteins were separated by SDS-PAGE transferred to polyvinylidene fluoride membranes (Immobilon P^{SEQ}, Millipore) using the Pharmacia Novablot transfer equipment and a continuous buffer system (39 mM glycine, 48 mM Tris, 0.0375% (w/v) SDS, 20% (v/v) methanol) according to the Multiphor II manual. The membranes were dried at 25°C for 1.5 h, blocked with 0.5% Tween-20 for 20 min, and then incubated with toxin A or toxin B antibodies (r-Biopharm, 0.2 µg/ml in TST buffer containing 0.05 M Tris, 0.5 M NaCl, 0.1% Tween-20, pH 9) for 1 h. After three washes in TST, the membranes were incubated with horse-radish peroxidase conjugated anti-mouse antibodies (DAKOPATTS, diluted 1:10000 in TST) for 1 h and washed three times in TST. A chemiluminiscent signal (ECL Plus, Amersham) was used to detect the bands. The relative amount of toxin B was measured on scanned x-ray films using the Molecular Analyst software (Biorad).

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE)

Protein samples were obtained as described under sample preparation. For 2-D PAGE, 40 ml aliquots of each sample was mixed with 160 ml of buffer III [9.9M Urea, 4% (v/v) Igepal CA630, 2.2% (v/v) ampholytes 3-10, 100mM DTT, 2% (w/v) CHAPS]. Protein mixtures were focused at 20°C on 180mm IPG Drystrip pH 4-7 (Pharmacia Biotech) using the Multiphor II 2-D gel Kit according to the manufacturerís instructions. The second dimension was run on 12% (linear) SDS acrylamide gels. The gels were stained with silver (PlusOne, Pharmacia Biotech) using a Hoeffer automatic gel stainer (Pharmacia Biotech). The chemicals were obtained from Sigma except for Pharmalytes (Pharmacia Biotech).

N-terminal sequencing

For N-terminal Sequence determination, gels were transferred to Immobilon-P polyvinyline difluoride membrane (Millipore) as described under immunoblotting. The membrane was stained with Coomassie blue; protein spots were excised for sequence determination. The protein spots cut from the transfer membrane were washed four times in 10% methanol and then dried and frozen. N-terminal sequence analysis was performed at the Protein Analysis Center, Karolinska Institute. Peptide sequences were matched against the *C.difficile* genome database (Sanger Center, UK).

References

Bardwell, J.C.A., Lee, J.-O., Jander, G., Martin, N., Belin, D., and Beckwith, J. (1993) A pathway for disulfide bond formation *in vivo*. *Proc Natl Acad Sci USA* **90**: 1038-1042.

Bardwell, J.C.A., McGovern, K., and Beckwith, J. (1991) Identification of a protein required for disulfide bond formation *in vivo*. *Cell* **65**: 581-589.

Boot, H.J., and Pouwels, P.H. (1996) Expression, secretion and antigenic variation of bacterial S-layer proteins. *Mol Microbiol* **21**: 1117-1123.

Callegari, L., Riboli, B., Sanders, P., Coconelli, S., Kok, J., Venema, G., and Morelli L.(1998) The S-layer gene of Lactobacillus helveticus CNRZ 892 : cloning, sequencing and heterologous expression. *Microbiology* **144**: 719-726.

Cerquetti, M., Pantosti, A., Stefanelli, P., and Mastrantonio, P. (1992) Purification and characterization of an immunodominant 36 kD antigen present on the cell surface of *Clostridium difficile*. *Microb Pathogen* **13**: 271-279.

Dauter, Z., Wilson, K.S., Sieker, L.C., Moulis, J.M., and Meyer, J. (1996) Zinc- and iron-rubredoxins from *Clostridium pasteurianum* at atomic resolution: a high-precision model for ZnS4 coordination unit in a protein. *Proc Natl Acad Sci USA* **93**: 8836-8840.

Dworkin, J., and Blaser, M.J. (1997) Molecular mechanisms of *Campylobacter fetus* surface layer protein expression. *Mol Microbiol* **26**: 433-440.

Economou, A. (1999) Following the leader : bacterial protein export through the Sec pathway. *Trends Microbiol* **7**: 315-319.

Etienne-Toumelin, I., Sirard, J.-C., Duflot, E., Mock, M., and Fouet, A. (1995) Characterization of the *Bacillus anthracis* S-layer : cloning and sequencing of the structural gene. *J. Bacteriol.* **177** : 614-620.

Garnier, T., and Cole, S.T. (1988) Studies of UV-inducible promoters from *Clostridium perfringens in vivo* and *in vitro*. *Mol Microbiol* **2**: 607-614.

Izard, J.W., and Kendall, D.A. (1994) Signal peptides: exquisitely designed transport promoters. *Mol Microbiol* **13**: 765-73.

Kamiya, S., Ogura, H., Meng, X.Q., and Nakamura, S. (1992) Correlation between cytotoxin production and sporulation in *Clostrdidium difficile*. *J Med Microbiol* **37**: 206-210.

Karasawa, T., Ikoma S., Yamakawa, K., and Nakamura, S. (1995) A defined medium for *Clostridium difficile*. *Microbiology* **141**: 371-375.

Kawata, T., Takeoka, A., Takumi, K., and Masuda, K. (1984) Demonstration and preliminary characterization of a regular array in the cell wall of *Clostridium difficile. FEMS Microbiol Lett* **24**: 232-328.

Klauser, T., Pohlner, J., and Meyer, T.F. (1993) The secretion pathway of IgA protease-type proteins in gram-negative bacteria. *Bioessays* **15**: 799-805.

Krogh, S., M. O'Reilly, N. Nolan, and K.M. Devine. (1996) The phage-like element PBSX and part of the skin element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. Microbiology **142** : 2031-2040.

Lazarevic, V., Margot, P., Suldo, B., and Karamata, D. (1992) Sequencing and analysis of the *Bacillus subtilis lytRABC* divergon : a regulatory unit encompassing the structural genes of the N-acetylmuromoyl-L-alanine amidase and its modifier. *J Gen Microbiol* **138**: 1949-1961.

Luckevich, M.D., and Beveridge, T.J. (1989) Characterization of a dynamic S-layer on *Bacillus thuringiensis. J. Bacteriol.* **171** : 6656-6667.

Masuda, K., Itoh, M., and Kawata, T. (1989) Characterization and reassembly of a regular array in the cell wall of *Clostridium difficile* GAI 4131. *Microbiol Immunol* **33**: 287-298.

Mauri, P.L., Pietta, P.G., Maggioni, A., Cerquetti, M., Sebastianelli, A., Mastrantonio, P. (1999) Characterization of surface layer proteins from *Clostridium difficile* by liquid chromatography/ electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **13**: 695-703.

McFarland, L.V., Elmer, G.W., Stamm, W.E., and Mulligan, M.E. 1991. Correlation of immunoblot type, enterotoxin production, and cytotoxin production with clinical manifestations of *Clostridium difficile* infection in a cohort of hospitalized patients. *Infect Immun* **59**: 2456-2462.

Mesnage, S., Tosi-Couture, E., Gounon, P., Mock, M., and Fouet, A. (1998) The capsule and the S-layer: two independent and yet compatible macromolecular structures in *Bacillus anthracis. J. Bacteriol.* **180** : 52-58.

Messner, P. (1996) Chemical composition and biosynthesis of S-layers, p.35-76. In U.B.Sleytr, P.Messner. D.Pum, and M.Sara (ed.), *Crystalline bacterial cell surface layer proteins (S-layers).* Academic Press, Landes Company, Austin, Tex.

Nicholls, L., Grant, T.H., and Robins-Browne, R.M. (2000) Identification of a novel genetic locus that is required for *in vitro* adhesion of a clinical isolate of enterohaemorrhagic *Escherichia coli* to epithelial cells. *Mol Microbiol* **35**: 275-288.

Orme-Johnson, W.H., and Orme-Johnson, N.R. (1978) Overview of iron-sulfur proteins. *Methods Enzymol* **53**: 259-268.

Pugsley, A.P. (1993) The complete general secretory pathway in Gram-negative bacteria. *Microbiol Rev* **57**: 50-108.

Sàra, M., and Sleytr, U.B. (2000) S-layer proteins. *J. Bacteriol.* **182** : 859-868.

Sàra, M., Kuen, B., Mayer, H.F., Mandl, F., Schuster, K.C., and Sleytr, U.B. (1996) Dynamics in oxygen induced changes in S-layer protein synthesis from *Bacillus stearothermophillus PV72* and the S-layer deficient variant T5 in continuous culture and studies on the cell wall composition. *J. Bacteriol.* **178** : 2108-2117.

Sellman, B.T., and Tweten, R.K. (1997) The propeptide of *Clostridium septicum* alpha toxin functions as an intramolecular chaperone and is a potent inhibitor of alpha toxin-dependent cytolysis. *Mol Microbiol* **25**: 429-440.

Simonen, M., and Palva, I. (1993) Protein secretion in *Bacillus* species. *Microbiol Rev* **57**: 109-137.

Takemaru, K., M. Mizuno, T. Sato, M. Takeuchi, and Y. Kobayashi. (1995) Complete nucleotide sequence of a skin element excised by DNA rearrangement during sporulation in *Bacillus subtilis*. Microbiology **141** : 323-327.

Takeoka, A., Takumi, K., Koga, T., and Kawata, T. (1991) Purification and characterization of S layer proteins from *Clostridium difficile* GAI 0714. *J Gen Microbiol* **137**: 261-267.

Takumi, K., Takeoka, A., and Kawata, T. (1987) Purification and immunochemical properties of a wall protein antigen from *Clostridium difficile* ATCC 110011. *Microbiol Immunol* **31**: 837-849.

Tsukagoshi, N., Tabata R., Takemura, T., Yamagata, H., and Udaka, S. (1984) Molecular cloning of a major cell wall protein gene from protein producing *Bacillus brevis 47* and its expression in *Escherichia coli* and *Bacillus subtilis*. *J. Bacteriol.* **158** : 1054-1060.

Welch, R.A. (1991) Pore-forming cytolysins of gram-negative bacteria. *Mol Microbiol* **5**: 21-528.

Table 1. Identification of ORF 1-12 and sequences upstream of ORF1 in SEQ ID NO: 35 and predicted signal peptide sequences of the S-layer ORFs. See Figures 1 and 2.

Gene	Frame	from	to	Total bp (aa)*	Homology/homologous motif (aa domain)*	Predicted signal peptide**	
A	+2	227	1366	1140	tRNA- Guanine transglycolase		
B	+3	1500	1793	294	conserved protein (membrane bound)		
C	+3	1851	2213	363	Hypothetical protein YrzE		
D	+1	3046	4941	1896 (631)	N-acetylmuramoyl-L-alanine amidase (aa 320-610)	MLSNKKRSMAIVMAGATVMSAAAPIFA↓DNTVTEN	SEQ ID NO: 8
E	+3	5373	6806	1434 (477)	N-acetylmuramoyl-L-alanine amidase (aa 20-290)	MKSTLGVENNMKNSKKILAIGLTLFLVMVNTPMV SA↓LTSVE	SEQ ID NO: 9
F	+2	7187	13150	5964	Unknown		
G	+1	13360	15240	1881 (626)	N-acetylmuramoyl-L-alanine amidase (aa 320-620)	MNKRKSFIRTIAVSTMAVAVTGSATCAYA↓APVLQ	SEQ ID NO: 10
H	+1	15556	17208	1653 (550)	N-acetylmuramoyl-L-alanine amidase (aa 5-370)	MENNHNINIKYKNHQGDMKMNKKILSLGLAVSLIL VNFKSVNA↓SSVV	SEQ ID NO: 11
I	+3	17427	19013	1587 (528)	N-acetylmuramoyl-L-alanine amidase (aa 20-250)	MKVNKRVLSIGLAISLIMAGAPNINA↓LSSIEK	SEQ ID NO: 12
ORF1	+3	19224	21383	2160 (719)	N-acetylmuramoyl-L-alanine amidase (aa 380-719)	MNKKNIAIAMSGLTVLASAAPVFA↓ATTGT	SEQ ID NO: 1
ORF2	+1	21628	23973	2346	SecA		SEQ ID NO: 34
ORF3	+1	24388	26259	1872 (622)	N-acetylmuramoyl-L-alanine amidase (aa 310-540)	MNKKNLSVIMAAAMISTSVAPVFA↓AETTQ	SEQ ID NO: 2
ORF4	+2	26420	27073	654	Unknown		
ORF5	+1	27106	28938	1833 (610)	N-acetylmuramoyl-L-alanine amidase (aa 5-300)	MKISKKIVSLLTMTFLTVTLYGNTSNA↓STKDT	SEQ ID NO: 3
ORF6	+3	29613	32024	2412 (803)	N-acetylmuramoyl-L-alanine amidase (aa 500-800)	MRKYKSKKLSKLLALSTVCFLIVSTIPVSA↓ENHK	SEQ ID NO: 4
ORF7	+2	32321	33898	578 (525)	N-acetylmuramoyl-L-alanine amidase (aa 30-340)	MKAPKTILTILTIALTLSSISIIPSYA↓LTEEK	SEQ ID NO: 5
ORF8	+3	34032	35351	1320	GDT1 protein		
ORF9	+1	35590	37629	2040 (528)	N-acetylmuramoyl-L-alanine amidase (aa 5-285)	MRGDMMKKTTKLLATGMLSVAMVAPNVALA↓AENTT	SEQ ID NO: 6
ORF10	+1	37873	38535	699	Hypothetical protein		
ORF11	+2	38630	39685	1056 (351)	N-acetylmuramoyl-L-alanine amidase (aa 40-340)	MIKKISTILSLVLLISISSTIGVFA↓DANPKR	SEQ ID NO: 7
ORF12	+2	39800	41356	1557	Virulence Factor MviN		

*aa = no. of amino acids or their position in translated ORF **↓ = predicted cleavage point

PCT/SE01/01280

Table 2. Possible promoter starts between ORF I and ORF 1. (1.00 is highest score). ORF 1 start codon is at position 19224.

Position	Score	Sequence	
19026-19071	0.91	TAGTTTATTACATTTTAAAATTTAGGGTATAAAAACTTGTAAACTTGGAG	SEQ ID NO: 13
19036-19081	1.00	CATTTTAAAATTTAGGGTATAAAAACTTGTAAACTTGGAGaAAATAATAA	SEQ ID NO: 14
19059-19104	0.80	AACTTGTAAACTTGGAGaAAATAATAATTTAAAAAAATAGCTTGCAAaAA	SEQ ID NO: 15
19067-19112	0.99	AACTTGGAGaAAATAATAATTTAAAAAAATAGCTTGCAAaAAGAATAAAA	SEQ ID NO: 16
19083-19128	1.00	TAATTTAAAAAAATAGCTTGCAAaAAGAATAAAAATGGATTATTATAGAG	SEQ ID NO: 17
19094-19139	0.99	AATAGCTTGCAAaAAGAATAAAAATGGATTATTATAGAGATGTGAGAaAT	SEQ ID NO: 18
19123-19168	0.96	TATTATAGAGATGTGAGAaATATTagGaATATATGGATGATTATTCTATG	SEQ ID NO: 19
19132-19177	0.85	GATGTGAGAaATATTagGaATATATGGATGATTATTCTATGtAcATAATA	SEQ ID NO: 20
19161-19206	0.84	GATTATTCTATGtAcATAATAAAGAGATGTAATTTTAATATAATGTTGGG	SEQ ID NO: 21

Table 3. Extracellular *C. difficile* proteins identified by N-terminal sequence analysis and genome database searches.

Spot	MW	Source	N-terminal Sequence[a]	Highest match[b]	Comment
1	50	PY	AAKASIADENSPVKLT LKSDXKKDL	No match	
2	36	PY	DDTKVETGDQGYTVV	N-acetyl---(ORF1)	Partly matched
3	47	PY	SEKEILTARLAV	Hypothetical protein	Fragment
4	40	PY	AIGLPSINISSK	PBSX phage protein, XkdK	
5	39	PY	MXDIKLDXFXKX	Fix B	
6	30	PY	MKILVXVKQVXX	Fix A	
7	41	PYG	AETTQVKKETIT	N-acetyl muramoyl-L-alanine amidase (ORF3)	Fragment
8	38	PYG	AETTQVKKETIT	N-acetyl---(ORF3)	Fragment
9	24	PYG	TSLKIADEVGLD	N-acetyl---(ORF1)	Fragment
10	48	PY	ANDTIASQDTPAKVV	N-acetyl---(ORF1)/ S-layer protein	Fragment (central part)
11	35	PY	ATTGTQGYTVVKNDD	N-acetyl---(ORF1)/S-layer protein	Fragment (N-terminal)

[a]Underlined sequence indicates that only this part was found to be identical with the ORF in the *C. difficile* strain 630 genome database

[b]Search was made in the *C. difficile* strain 630 genome database by tBLASTn algorithm at www.. Contigs were exported to ORF finder at NCBI, and the entire ORF was subjected to further BLAST searches against the redundant database.

Table 4. Blast search summary of the ORFs in the *C. difficile* strain 630 S- layer contig. ORF numbers refer to those in Fig 1.

ORF	Length (amino acids)	Match (highest E-value)[1]	Source	E value
1	719	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$1x10^{-23}$
2	781	Preprotein translocase SecA subunit	*B. subtilis*	0.0
3	623	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$1x10^{-24}$
4	217	No match		
5	610	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$2x10^{-42}$
6	803	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$2x10^{-36}$
7	525	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$2x10^{-39}$
8	439	GDT1 protein	*D. discoideum*	$2x10^{-11}$
9	679	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$4x10^{-56}$
10	232	Hypothetical protein	*B. subtilis*	$7x10^{-45}$
11	351	N-acetylmuramoyl-L-alanine amidase	*B. subtilis*	$1x10^{-11}$
12	518	Virulence factor MviN	*S. typhimurium*	$2x10^{-53}$
13	568	Phosphomannose murase	*B. subtilis*	10^{-159}
14	352	Mannose-1-phosphate guanyltransferase	*E. coli*	$5x10^{-68}$
15	157	Cap8J	*S. pnemoniae*	$5x10^{-6}$
16	385	No match		
17	279	Glucosyl transferase	*S. pnemoniae*	$5x1^{-11}$

[1]Database search was made using the BLAST algorithm at http://www.ncbi.nlm.nih.gov/.com.

Matches with a higher E value than 0.001 was regarded as no match.

CLAIMS

1. A gene expression cassette comprising a secretory leader sequence encoding a signal peptide from *Clostridium difficile* having an amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7 and signal peptides of analogous exported clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked to a DNA sequence encoding a heterologous polypeptide.

2. A gene expression cassette according to claim 1, wherein the signal peptides of the analogous clostridial N-acetylmuramoyl-L-alanine amidase-like proteins are selected from *Clostridium difficile* signal peptides having an amino acid sequence of any one of SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12.

3. A gene expression cassette according to claim 1 or 2, which further includes a promoter of prokaryotic origin.

4. A gene expression cassette according to claim 3, wherein the promoter is selected from clostridial promoters comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 13 – 21.

5. A gene expression cassette according to any one of claims 1-4 which further includes a DNA sequence encoding at least a cell wall binding portion of a protein of prokaryotic origin functioning in clostridia such that a fusion polypeptide may be presented on the outer surface of a host cell harbouring the cassette.

6. A gene expression cassette according to any one of claims 1-4, which further includes a DNA sequence encoding at least a functional cell wall binding portion of an S-layer protein of *C. difficile* selected from any one of the polypeptides having an amino acid sequence selected from SEQ ID NO: 22 - 33 such that a fusion polypeptide may be presented on the outer surface of a host cell harbouring the cassette.

7. A gene expression cassette according to claim 5 or 6, wherein DNA encoding the cell wall binding portions of SEQ ID NO: 22-33 has been omitted such that the fusion peptide is secreted into the surrounding milieu by the host cell harbouring the cassette.

8. A gene expression cassette according to claim 6 or 7, wherein the DNA sequence encoding the heterologous peptide is inserted at a point downstream the first (signal) proteolytic cleavage site in the gene encoding a polypeptide having an amino acid sequence selected from SEQ ID NO: 22 – 33, optionally including or excluding its second cleavage site.

9. A gene expression cassette according to any one of claims 1-8, which further comprises at least a functional part of a secretory (*secA*) gene recognizing the signal peptide, to allow translocation of a heterologous polypeptide and/or fusion polypeptide across the cytoplasmic membrane of a host cell harbouring the expression cassette.

10. A gene expression cassette according to claim 9, wherein the secretory gene is from *C. difficile* and encodes a polypeptide having the amino acid sequence SEQ ID NO: 34.

11. A gene expression cassette as shown in Figure 3.

12. A vector comprising a gene expression cassette as claimed in any one of claims 1-11.

13. A vector according to claim 12, wherein the vector is a plasmid.

14. A host organism transformed with a vector according to claim 12 or 13 for expression of the heterologous polypeptide and/or fusion polypeptide.

15. A *Clostridium* host organism transformed with a vector according to claim 12 or 13 for expression of the heterologous polypeptide and/or fusion polypeptide.

16. A host organism as claimed in claim 15 which is *C. difficile*.

17. A host organism as claimed in claim 15 which is *C. perfringens*.

18. A pharmaceutical or veterinary composition or formulation which comprises *Clostridial* cells transformed with a vector according to claim 12 or 13, with the ability to present on the cell surface and/or to secrete an expressed heterologous polypeptide or fusion polypeptide, together with a pharmaceutically or veterinary acceptable carrier or diluent.

19. A composition or formulation according to claim 18, which is suitable for oral or intranasal administration.

20. A composition or formulation according to claim 18 or 19, which further comprises, as an adjuvant, non-toxic motifs of the *C. difficile* toxins A and/or B that enable the expressed heterologous polypeptide and/or fusion polypeptide to enter the colonic mucosal cells of a mammal by receptor-mediated endocytosis, and/or a portion of toxin B responsible for its intracellular and intercellular spread.

21. A composition or formulation according to claim 18 or 19, which further comprises a further fused or separate carrier peptide or adjuvant, in addition to the expressed heterologous polypeptide and/or fusion polypeptide, to elicit a stronger or differently directed immune response than that against the expressed heterologous polypeptide acting alone.

22. A vaccine which comprises a *Clostridial* bacterium transformed with a vector according to claim 12 or 13 and capable of presenting on the surface of the bacterium and/or secreting an antigen in a human or animal body, and optionally also an adjuvant described in claim 20 or 21.

23. A medicament which comprises a *Clostridial* bacterium transformed with a vector according to claim 12 or 13 and capable of presenting on the surface of the bacterium and/or secreting a heterologous polypeptide and/or fusion polypeptide in a human or animal body, and optionally also an adjuvant described in claim 20 or 21.

24. A vaccine according to claim 22, which comprises a mixture of at least two differently engineered *Clostridia* strains, each capable of presenting on the surface of the bacterium and/or secreting a different heterologous polypeptide and/or fusion polypeptide.

25. A medicament according to claim 23, which comprises a mixture of at least two differently engineered *Clostridia* strains, each capable of presenting on the surface of the bacterium and/or secreting a different heterologous polypeptide and/or fusion polypeptide.

26. A vaccine which comprises spores of *Clostridia* cells or bacteria transformed with a vector according to claim 12 or 13, and capable of germinating into cells which are able to grow, express, and present or secrete a heterologous polypeptide and/or fusion polypeptide, and optionally an adjuvant described in claim 20 or 21, in a mammalian body.

27. A medicament which comprises spores of *Clostridia* cells or bacteria transformed with a vector according to claim 12 or 13, and capable of germinating into cells which are able to grow, express, and present or secrete a heterologous polypeptide and/or fusion polypeptide and optionally an adjuvant described in claim 20 or 21, in a mammalian body.

28. A vaccine according to claim 26, which comprises a mixture of spores from at least two differently engineered *Clostridia* strains.

29. A medicament according to claim 27, which comprises a mixture of spores from at least two differently engineered *Clostridia* strains.

30. A method for vaccination of a mammal, which comprises administering a therapeutically or prophylactically effective dose of a vaccine according to any one of claims 22, 24, 26 and 28 to the mammal.

31. A method for prophylactic or therapeutic treatment of a mammal, which comprises administering a therapeutically or prophylactically effective dose of a medicament according to any one of claims 23, 25, 27 and 29 to the mammal.

32. A vaccine according to claim 26 or 28, a medicament according to claim 27 or 29, or a method according to claim 30 or 31, wherein the spores are from *C. difficile* or *C. perfringens*.

33. A *C. difficile*-associated diarrhea (CDAD) vaccine comprising spores according to claim 26 or 32 and capable of expressing, after germination,

(i) relevant parts of the *C. difficile* toxins, alone or together with a

(ii) suitable adjuvant to provide primarily an IgA response to the toxin antigenic epitopes and/or

(iii) S-layer protein antigenic variants (serotype antigens) or fimbrial antigens from *C. difficile* to obtain, after administration to a mammal, a polyvalent anti-S-layer (or anti-fimbrial) IgA response preventing *C. difficile* colonization of the mammal.

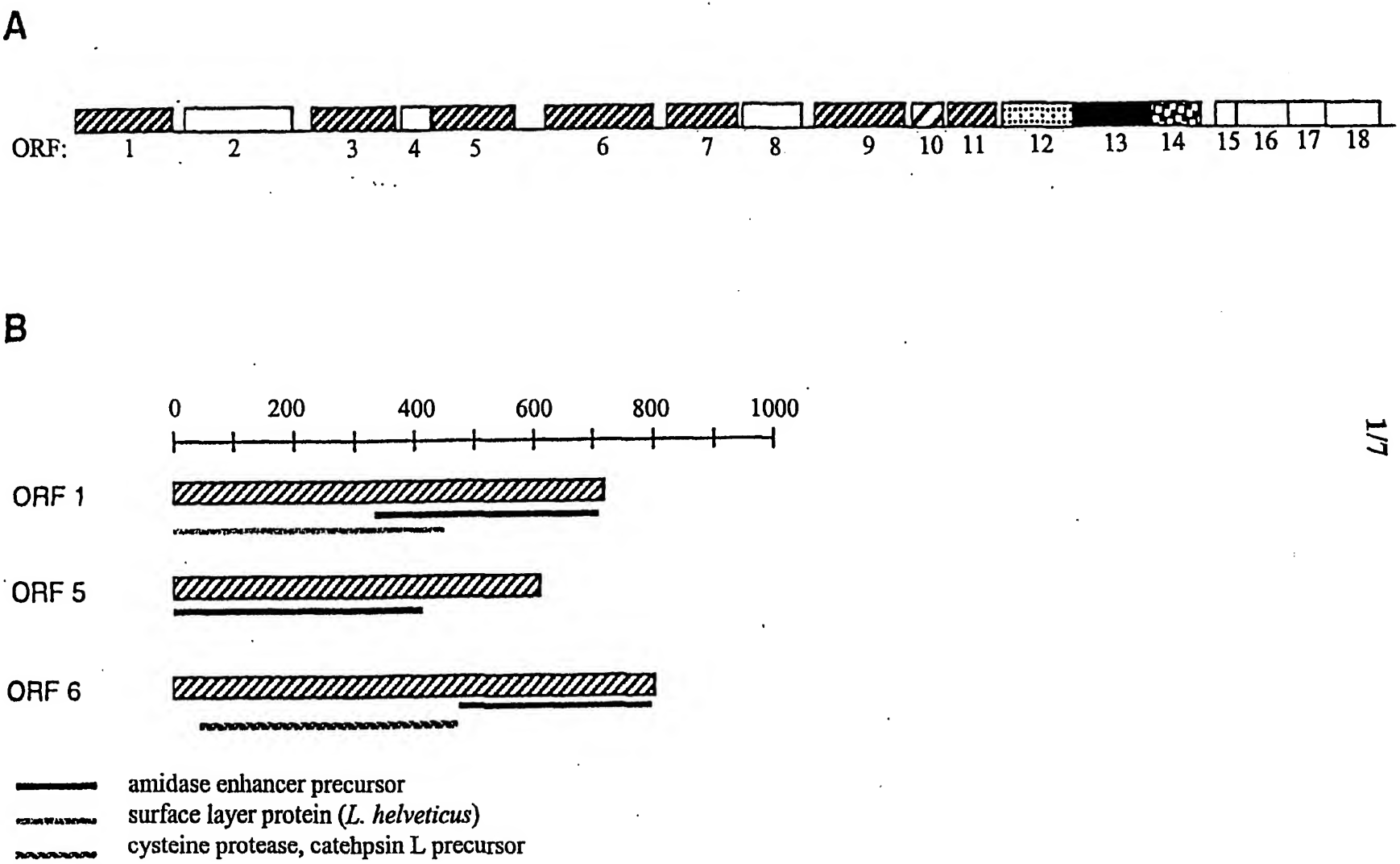

A
ORF: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
B
0 200 400 600 800 1000
ORF 1
ORF 5
ORF 6
amidase enhancer precursor
surface layer protein (L. helveticus)
cysteine protease, catehpsin L precursor

Figure 1

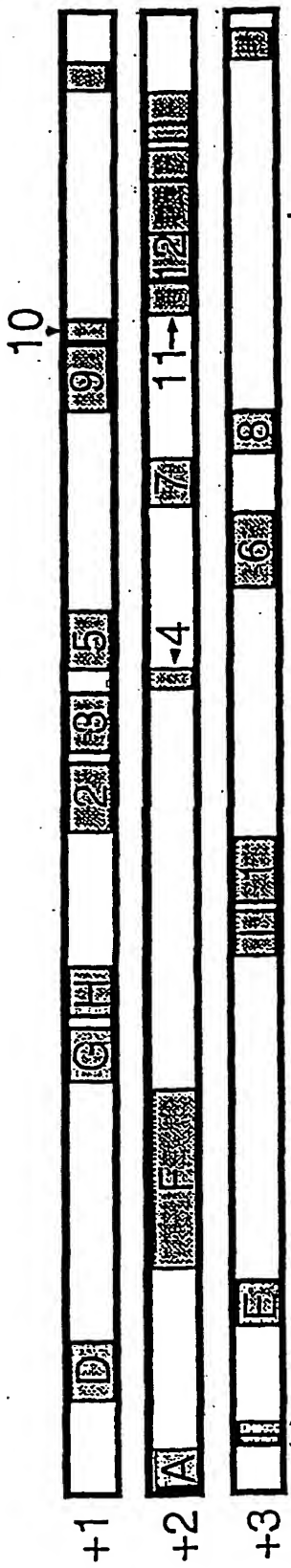

10
D
G
H
2
3
5
9
E
A
F
4
7
11
12
1
6
8
B
C
+1
+2
+3
ORFs 1 to 12
Upstream region
ORFs A-I

Figure 2

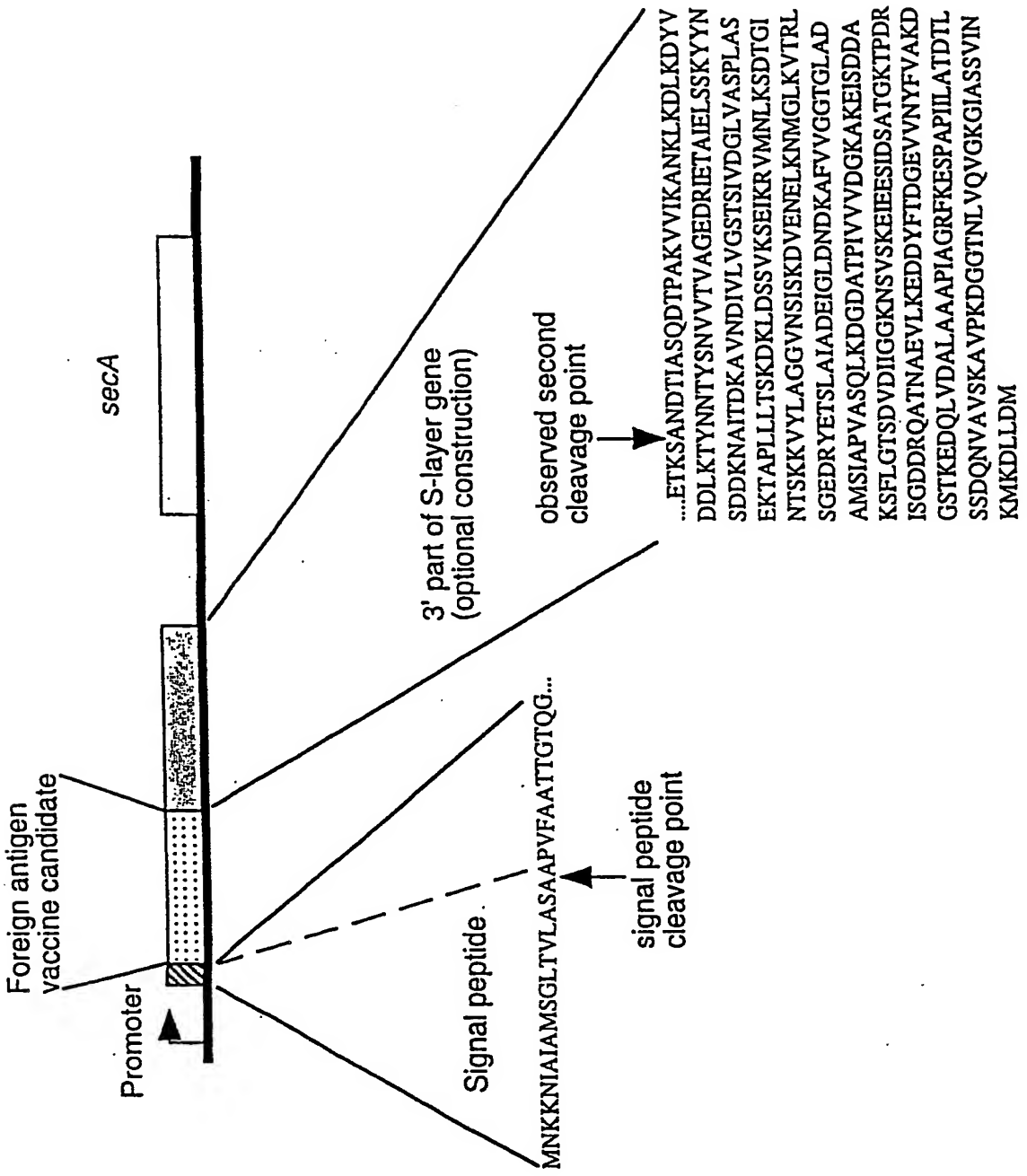

Foreign antigen vaccine candidate
Promoter
secA
Signal peptide
MNKKNIAIAMSGLTVLASAAPVFAATTGTQG...
signal peptide cleavage point
3' part of S-layer gene (optional construction)
observed second cleavage point
...ETKSANDTIASQDTPAKVVIKANKLKDLKDYV
DDLKTYNNTYSNVVTVAGEDRIETAIELSSKYYN
SDDKNAITDKAVNDIVLVGSTSIVDGLVASPLAS
EKTAPLLLTSKDKLDSSVKSEIKRVMNLKSDTGI
NTSKKVYLAGGVNSISKDVENELKNMGLKVTRL
SGEDRYETSLAIADEIGLDNDKAFVVGGTGLAD
AMSIAPVASQLKDGDATPIVVVDGKAKEISDDA
KSFLGTSDVDIIGGKNSVSKEIEESIDSATGKTPDR
ISGDDRQATNAEVLKEDDYFTDGEVVNYFVAKD
GSTKEDQLVDALAAAPIAGRFKESPAPIILATDTL
SSDQNVAVSKAVPKDGGTNLVQVGKGIASSVIN
KMKDLLDM

Figure 3

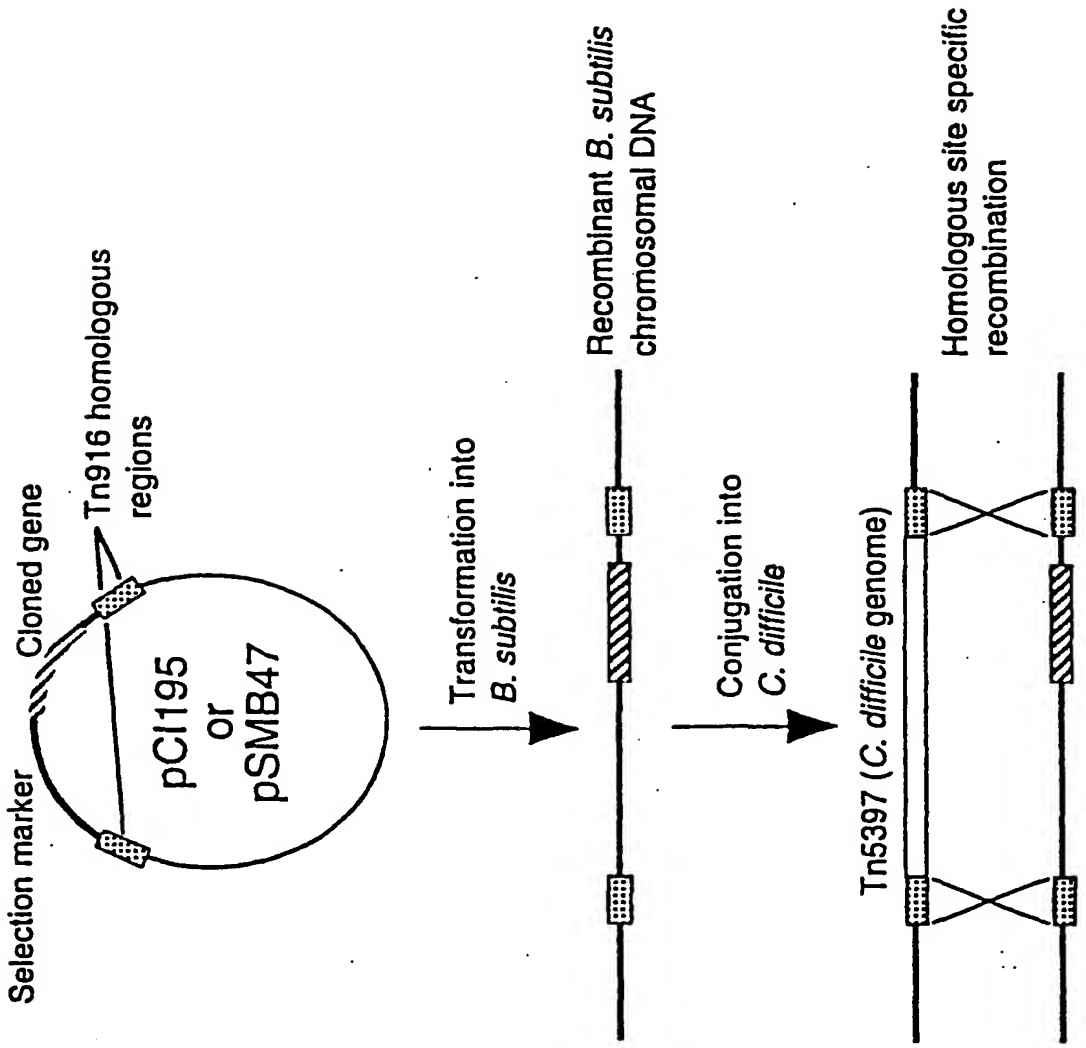

Selection marker
Cloned gene
Tn916 homologous regions
pCl195 or pSMB47
Transformation into B. subtilis
Recombinant B. subtilis chromosomal DNA
Conjugation into C. difficile
Tn5397 (C. difficile genome)
Homologous site specific recombination

Figure 4

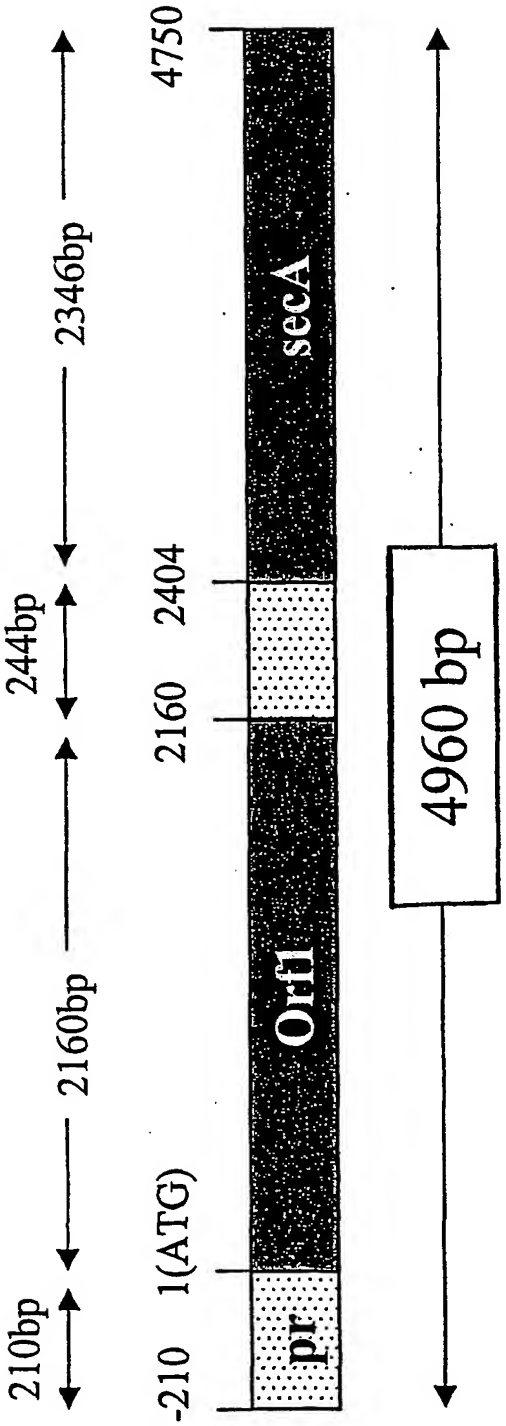

4750
secA
2346bp
2404
244bp
2160
Orf1
2160bp
1(ATG)
4960 bp
210bp
-210
pr

Figure 5

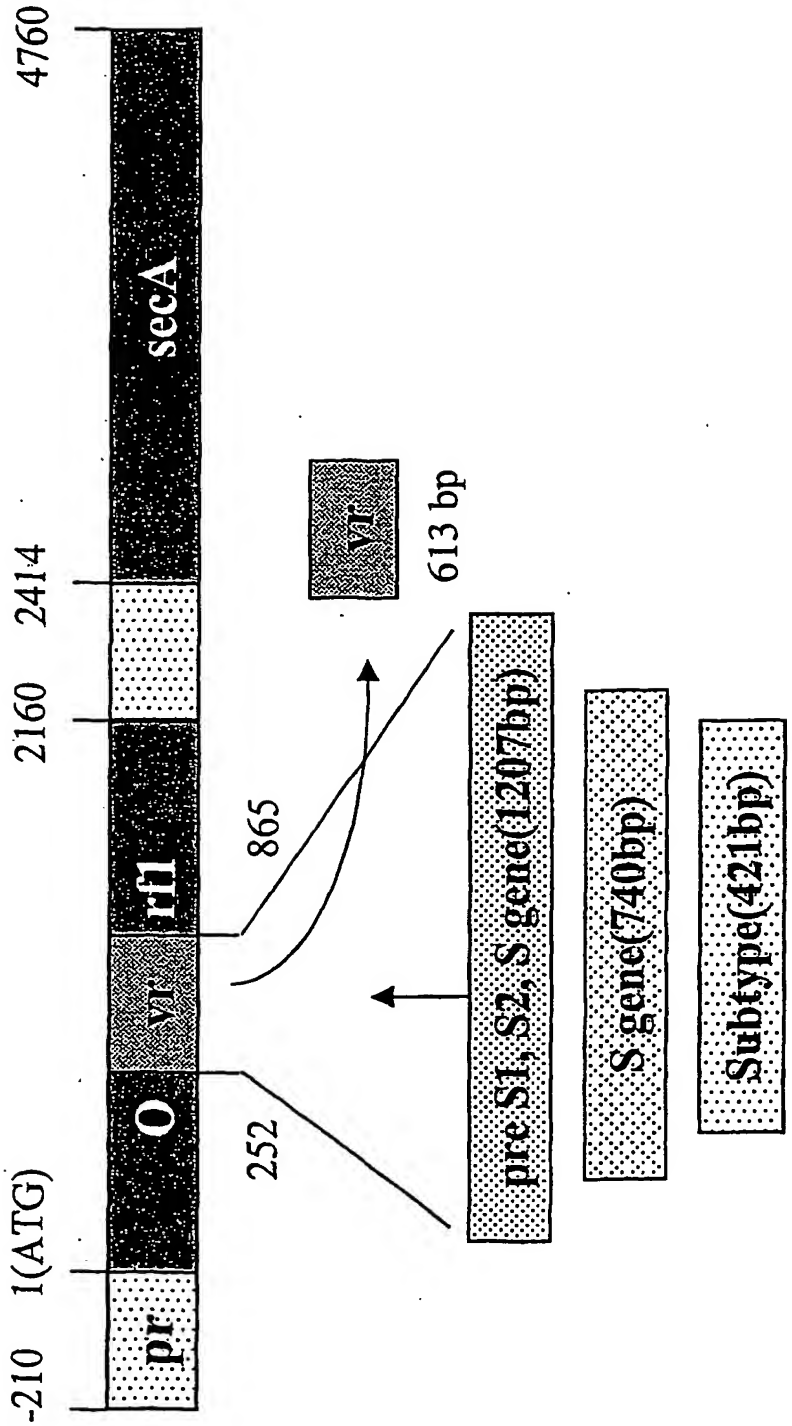
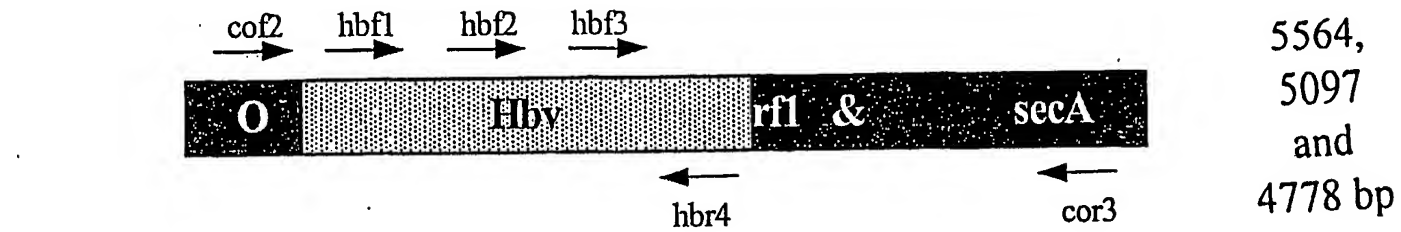

4760
2414
2160
1(ATG)
-210
secA
rt1
VL
pr
O
865
252
VL
613 bp
pre S1, S2, S gene(1207bp)
S gene(740bp)
Subtype(421bp)

Figure 6


cof2
hbf1
hbf2
hbf3
O
Hby
rf1 & secA
hbr4
cor3
5564,
5097
and
4778 bp
Primers cof2 and cor3 designed to amplify orf1 & secA with native promoter
Primers hbf1, 2,3 and hbr4 designed to amplify Hbv

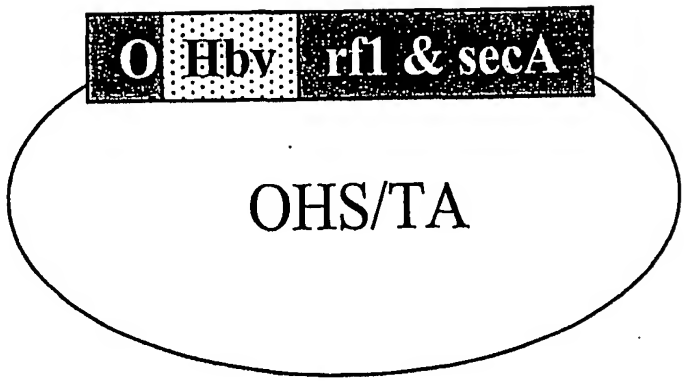

O
Hby
rf1 & secA
OHS/TA

Figure 7

SEQUENCE LISTING

<110> Smittskyddsinstitutet

<120> Gene expression cassette and its use

<130> 110060300

<140>
<141>

<150> SE 0002139-4
<151> 2000-06-07

<150> 0101479-4
<151> 2001-04-26

<160> 35

<170> PatentIn Ver. 2.1

<210> 1
<211> 29
<212> PRT
<213> Clostridium difficile

<400> 1
Met Asn Lys Lys Asn Ile Ala Ile Ala Met Ser Gly Leu Thr Val Leu
1 5 10 15

Ala Ser Ala Ala Pro Val Phe Ala Ala Thr Thr Gly Thr
20 25

<210> 2
<211> 29
<212> PRT
<213> Clostridium difficile

<400> 2
Met Asn Lys Lys Asn Leu Ser Val Ile Met Ala Ala Ala Met Ile Ser
1 5 10 15

Thr Ser Val Ala Pro Val Phe Ala Ala Glu Thr Thr Gln
20 25

<210> 3
<211> 32
<212> PRT
<213> Clostridium difficile

<400> 3
Met Lys Ile Ser Lys Lys Ile Val Ser Leu Leu Thr Met Thr Phe Leu
1 5 10 15

Thr Val Thr Leu Tyr Gly Asn Thr Ser Asn Ala Ser Thr Lys Asp Thr
20 25 30

<210> 4
<211> 34
<212> PRT
<213> Clostridium difficile

<400> 4
Met Arg Lys Tyr Lys Ser Lys Lys Leu Ser Lys Leu Leu Ala Leu Ser
1 5 10 15

Thr Val Cys Phe Leu Ile Val Ser Thr Ile Pro Val Ser Ala Glu Asn
20 25 30

His Lys

<210> 5
<211> 32
<212> PRT
<213> Clostridium difficile

<400> 5
Met Lys Ala Pro Lys Thr Ile Leu Thr Ile Leu Thr Ile Ala Leu Thr
1 5 10 15

Leu Ser Ser Ile Ser Ile Ile Pro Ser Tyr Ala Leu Thr Glu Glu Lys
20 25 30

<210> 6
<211> 35
<212> PRT
<213> Clostridium difficile

<400> 6
Met Arg Gly Asp Met Met Lys Lys Thr Thr Lys Leu Leu Ala Thr Gly
1 5 10 15

Met Leu Ser Val Ala Met Val Ala Pro Asn Val Ala Leu Ala Ala Glu
20 25 30

Asn Thr Thr
35

<210> 7
<211> 31
<212> PRT
<213> Clostridium difficile

<400> 7
Met Ile Lys Lys Ile Ser Thr Ile Leu Ser Leu Val Leu Leu Ile Ser
1 5 10 15

Ile Ser Ser Thr Ile Gly Val Phe Ala Asp Ala Asn Pro Lys Arg
20 25 30

<210> 8
<211> 34
<212> PRT
<213> Clostridium difficile

<400> 8
Met Leu Ser Asn Lys Lys Arg Ser Met Ala Ile Val Met Ala Gly Ala
1 5 10 15

Thr Val Met Ser Ala Ala Ala Pro Ile Phe Ala Asp Asn Thr Val Thr
20 25 30

Glu Asn

<210> 9
<211> 41
<212> PRT
<213> Clostridium difficile

<400> 9
Met Lys Ser Thr Leu Gly Val Glu Asn Asn Met Lys Asn Ser Lys Lys
1 5 10 15

Ile Leu Ala Ile Gly Leu Thr Leu Phe Leu Val Met Val Asn Thr Pro
20 25 30

Met Val Ser Ala Leu Thr Ser Val Glu
35 40

<210> 10
<211> 34
<212> PRT
<213> Clostridium difficile

<400> 10
Met Asn Lys Arg Lys Ser Phe Ile Arg Thr Ile Ala Val Ser Thr Met
1 5 10 15

Ala Val Ala Val Thr Gly Ser Ala Thr Cys Ala Tyr Ala Ala Pro Val
20 25 30

Leu Gln

<210> 11
<211> 47
<212> PRT
<213> Clostridium difficile

<400> 11
Met Glu Asn Asn His Asn Ile Asn Ile Lys Tyr Lys Asn His Gln Gly
1 5 10 15

Asp Met Lys Met Asn Lys Lys Ile Leu Ser Leu Gly Leu Ala Val Ser
20 25 30

Leu Ile Leu Val Asn Phe Lys Ser Val Asn Ala Ser Ser Val Val
35 40 45

<210> 12
<211> 32
<212> PRT
<213> Clostridium difficile

<400> 12
Met Lys Val Asn Lys Arg Val Leu Ser Ile Gly Leu Ala Ile Ser Leu
1 5 10 15

Ile Met Ala Gly Ala Pro Asn Ile Asn Ala Leu Ser Ser Ile Glu Lys
20 25 30

<210> 13
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 13
tagtttatta cattttaaaa tttagggtat aaaaacttgt aaacttggag 50

<210> 14
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 14
cattttaaaa tttagggtat aaaaacttgt aaacttggag aaaataataa 50

<210> 15
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 15
aacttgtaaa cttggagaaa ataataattt aaaaaaatag cttgcaaaaa 50

<210> 16
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 16
aacttggaga aaataataat ttaaaaaaat agcttgcaaa aagaataaaa 50

<210> 17
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 17
taatttaaaa aaatagcttg caaaaagaat aaaaatggat tattatagag 50

<210> 18
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 18
aatagcttgc aaaaagaata aaaatggatt attatagaga tgtgagaaat 50

<210> 19
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 19
tattatagag atgtgagaaa tattaggaat atatggatga ttattctatg 50

<210> 20
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 20
gatgtgagaa atattaggaa tatatggatg attattctat gtacataata 50

<210> 21
<211> 50
<212> DNA
<213> Clostridium difficile

<400> 21
gattattcta tgtacataat aaagagatgt aattttaata taatgttggg 50

<210> 22
<211> 719
<212> PRT
<213> Clostridium difficile

<400> 22
Met Asn Lys Lys Asn Ile Ala Ile Ala Met Ser Gly Leu Thr Val Leu
1 5 10 15

Ala Ser Ala Ala Pro Val Phe Ala Ala Thr Thr Gly Thr Gln Gly Tyr
20 25 30

Thr Val Val Lys Asn Asp Trp Lys Lys Ala Val Lys Gln Leu Gln Asp
35 40 45

Gly Leu Lys Asp Asn Ser Ile Gly Lys Ile Thr Val Ser Phe Asn Asp
50 55 60

Gly Val Val Gly Glu Val Ala Pro Lys Ser Ala Asn Lys Lys Ala Asp
65 70 75 80

Arg Asp Ala Ala Ala Glu Lys Leu Tyr Asn Leu Val Asn Thr Gln Leu
85 90 95

```
Asp Lys Leu Gly Asp Gly Asp Tyr Val Asp Phe Ser Val Asp Tyr Asn
            100                 105                 110

Leu Glu Asn Lys Ile Ile Thr Asn Gln Ala Asp Ala Glu Ala Ile Val
        115                 120                 125

Thr Lys Leu Asn Ser Leu Asn Glu Lys Thr Leu Ile Asp Ile Ala Thr
    130                 135                 140

Lys Asp Thr Phe Gly Met Val Ser Lys Thr Gln Asp Ser Glu Gly Lys
145                 150                 155                 160

Asn Val Ala Ala Thr Lys Ala Leu Lys Val Lys Asp Val Ala Thr Phe
                165                 170                 175

Gly Leu Lys Ser Gly Gly Ser Glu Asp Thr Gly Tyr Val Val Glu Met
            180                 185                 190

Lys Ala Gly Ala Val Glu Asp Lys Tyr Gly Lys Val Gly Asp Ser Thr
        195                 200                 205

Ala Gly Ile Ala Ile Asn Leu Pro Ser Thr Gly Leu Glu Tyr Ala Gly
    210                 215                 220

Lys Gly Thr Thr Ile Asp Phe Asn Lys Thr Leu Lys Val Asp Val Thr
225                 230                 235                 240

Gly Gly Ser Thr Pro Ser Ala Val Ala Val Ser Gly Phe Val Thr Lys
                245                 250                 255

Asp Asp Thr Asp Leu Ala Lys Ser Gly Thr Ile Asn Val Arg Val Ile
            260                 265                 270

Asn Ala Lys Glu Glu Ser Ile Asp Ile Asp Ala Ser Ser Tyr Thr Ser
        275                 280                 285

Ala Glu Asn Leu Ala Lys Arg Tyr Val Phe Asp Pro Asp Glu Ile Ser
    290                 295                 300

Glu Ala Tyr Lys Ala Ile Val Ala Leu Gln Asn Asp Gly Ile Glu Ser
305                 310                 315                 320

Asn Leu Val Gln Leu Val Asn Gly Lys Tyr Gln Val Ile Phe Tyr Pro
                325                 330                 335

Glu Gly Lys Arg Leu Glu Thr Lys Ser Ala Asn Asp Thr Ile Ala Ser
            340                 345                 350

Gln Asp Thr Pro Ala Lys Val Val Ile Lys Ala Asn Lys Leu Lys Asp
        355                 360                 365

Leu Lys Asp Tyr Val Asp Asp Leu Lys Thr Tyr Asn Asn Thr Tyr Ser
    370                 375                 380

Asn Val Val Thr Val Ala Gly Glu Asp Arg Ile Glu Thr Ala Ile Glu
385                 390                 395                 400

Leu Ser Ser Lys Tyr Tyr Asn Ser Asp Asp Lys Asn Ala Ile Thr Asp
                405                 410                 415
```

```
Lys Ala Val Asn Asp Ile Val Leu Val Gly Ser Thr Ser Ile Val Asp
            420                 425                 430

Gly Leu Val Ala Ser Pro Leu Ala Ser Glu Lys Thr Ala Pro Leu Leu
        435                 440                 445

Leu Thr Ser Lys Asp Lys Leu Asp Ser Ser Val Lys Ser Glu Ile Lys
    450                 455                 460

Arg Val Met Asn Leu Lys Ser Asp Thr Gly Ile Asn Thr Ser Lys Lys
465                 470                 475                 480

Val Tyr Leu Ala Gly Gly Val Asn Ser Ile Ser Lys Asp Val Glu Asn
                485                 490                 495

Glu Leu Lys Asn Met Gly Leu Lys Val Thr Arg Leu Ser Gly Glu Asp
            500                 505                 510

Arg Tyr Glu Thr Ser Leu Ala Ile Ala Asp Glu Ile Gly Leu Asp Asn
        515                 520                 525

Asp Lys Ala Phe Val Val Gly Gly Thr Gly Leu Ala Asp Ala Met Ser
    530                 535                 540

Ile Ala Pro Val Ala Ser Gln Leu Lys Asp Gly Asp Ala Thr Pro Ile
545                 550                 555                 560

Val Val Val Asp Gly Lys Ala Lys Glu Ile Ser Asp Asp Ala Lys Ser
                565                 570                 575

Phe Leu Gly Thr Ser Asp Val Asp Ile Ile Gly Gly Lys Asn Ser Val
            580                 585                 590

Ser Lys Glu Ile Glu Glu Ser Ile Asp Ser Ala Thr Gly Lys Thr Pro
        595                 600                 605

Asp Arg Ile Ser Gly Asp Asp Arg Gln Ala Thr Asn Ala Glu Val Leu
    610                 615                 620

Lys Glu Asp Asp Tyr Phe Thr Asp Gly Glu Val Val Asn Tyr Phe Val
625                 630                 635                 640

Ala Lys Asp Gly Ser Thr Lys Glu Asp Gln Leu Val Asp Ala Leu Ala
                645                 650                 655

Ala Ala Pro Ile Ala Gly Arg Phe Lys Glu Ser Pro Ala Pro Ile Ile
            660                 665                 670

Leu Ala Thr Asp Thr Leu Ser Ser Asp Gln Asn Val Ala Val Ser Lys
        675                 680                 685

Ala Val Pro Lys Asp Gly Gly Thr Asn Leu Val Gln Val Gly Lys Gly
    690                 695                 700

Ile Ala Ser Ser Val Ile Asn Lys Met Lys Asp Leu Leu Asp Met
705                 710                 715
```

<210> 23
<211> 623
<212> PRT

<213> Clostridium difficile

<400> 23

```
Met Asn Lys Lys Asn Leu Ser Val Ile Met Ala Ala Ala Met Ile Ser
  1               5                  10                  15

Thr Ser Val Ala Pro Val Phe Ala Ala Glu Thr Thr Gln Val Lys Lys
             20                  25                  30

Glu Thr Ile Thr Lys Lys Glu Ala Thr Glu Leu Val Ser Lys Val Arg
         35                  40                  45

Asp Leu Met Ser Gln Lys Tyr Thr Gly Gly Ser Gln Val Gly Gln Pro
     50                  55                  60

Ile Tyr Glu Ile Lys Val Gly Glu Thr Leu Ser Lys Leu Lys Ile Ile
 65                  70                  75                  80

Thr Asn Ile Asp Glu Leu Glu Lys Leu Val Asn Ala Leu Gly Glu Asn
                 85                  90                  95

Lys Glu Leu Ile Val Thr Ile Thr Asp Lys Gly His Ile Thr Asn Ser
            100                 105                 110

Ala Asn Glu Val Val Ala Glu Ala Thr Glu Lys Tyr Glu Asn Ser Ala
        115                 120                 125

Asp Leu Ser Ala Glu Ala Asn Ser Ile Thr Glu Lys Ala Lys Thr Glu
    130                 135                 140

Thr Asn Gly Ile Tyr Lys Val Ala Asp Val Lys Ala Ser Tyr Asp Ser
145                 150                 155                 160

Ala Lys Asp Lys Leu Val Ile Thr Leu Arg Asp Lys Thr Glu Thr Val
                165                 170                 175

Thr Ser Lys Thr Ile Tyr Val Gly Ile Gly Asp Glu Lys Val Asp Leu
            180                 185                 190

Thr Ala Asn Pro Val Asp Ser Thr Gly Thr Asn Leu Asp Pro Ser Ala
        195                 200                 205

Glu Gly Phe Arg Val Asn Lys Ile Asp Lys Leu Gly Val Ala Gly Ala
    210                 215                 220

Lys Asn Ile Asp Asp Val Gln Leu Ala Glu Ile Thr Ile Lys Asn Ser
225                 230                 235                 240

Asp Leu Asn Thr Val Ser Pro Gln Asp Leu Tyr Asp Gly Tyr Arg Leu
                245                 250                 255

Thr Val Lys Gly Asn Met Val Ala Asn Gly Thr Ser Lys Ser Ile Ser
            260                 265                 270

Asp Ile Ser Ala Lys Asp Ser Glu Thr Gly Lys Tyr Lys Phe Thr Ile
        275                 280                 285

Lys Tyr Thr Asp Ala Ser Gly Lys Ala Thr Glu Leu Thr Val Glu Ser
    290                 295                 300
```

```
Thr Asn Glu Lys Asp Leu Lys Asp Ala Lys Ala Ala Leu Glu Gly Asn
305                 310                 315                 320

Ser Lys Val Lys Leu Ile Ala Gly Asp Asp Arg Tyr Ala Thr Ala Val
                325                 330                 335

Ala Ile Ala Lys Gln Thr Lys Tyr Thr Asp Asn Ile Val Ile Val Asn
            340                 345                 350

Ser Asn Lys Leu Val Asp Gly Leu Ala Ala Thr Pro Leu Ala Gln Ser
        355                 360                 365

Lys Lys Ala Pro Ile Leu Leu Ala Ser Asp Asn Glu Ile Pro Lys Val
    370                 375                 380

Thr Leu Asp Tyr Ile Lys Asp Ile Ile Lys Lys Ser Pro Ser Ala Lys
385                 390                 395                 400

Ile Tyr Ile Val Gly Gly Glu Ser Ala Val Ser Asn Thr Ala Lys Lys
                405                 410                 415

Gln Leu Glu Ser Val Thr Lys Asn Ile Glu Arg Leu Ala Gly Asp Asp
            420                 425                 430

Arg His Thr Thr Ser Val Ala Val Ala Lys Ala Met Gly Ser Phe Lys
        435                 440                 445

Asp Ala Phe Val Val Gly Ala Lys Gly Glu Ala Asp Ala Met Ser Ile
    450                 455                 460

Ala Ala Lys Ala Ala Glu Leu Lys Ala Pro Ile Ile Val Asn Gly Trp
465                 470                 475                 480

Asn Asp Leu Ser Ala Asp Ala Ile Lys Leu Met Asp Gly Lys Glu Ile
                485                 490                 495

Gly Ile Val Gly Gly Ser Asn Asn Val Ser Ser Gln Ile Glu Asn Gln
            500                 505                 510

Leu Ala Asp Ile Asp Lys Asp Arg Lys Val Gln Arg Val Glu Gly Glu
        515                 520                 525

Thr Arg His Asp Thr Asn Ala Lys Val Ile Glu Thr Tyr Tyr Gly Lys
    530                 535                 540

Leu Asp Lys Leu Tyr Ile Ala Lys Asp Gly Tyr Gly Asn Asn Gly Met
545                 550                 555                 560

Leu Val Asp Ala Leu Ala Ala Gly Pro Leu Ala Ala Gly Lys Gly Pro
                565                 570                 575

Ile Leu Leu Ala Lys Thr Asp Ile Thr Asp Ser Gln Lys Asn Ala Leu
            580                 585                 590

Ser Lys Lys Leu Asn Leu Gly Ala Glu Val Thr Gln Ile Gly Asn Gly
        595                 600                 605

Val Glu Leu Thr Val Ile Gln Lys Ile Ala Lys Ile Leu Gly Trp
    610                 615                 620
```

<210> 24

<211> 610
<212> PRT
<213> Clostridium difficile

<400> 24

```
Met Lys Ile Ser Lys Lys Ile Val Ser Leu Leu Thr Met Thr Phe Leu
  1               5                  10                  15

Thr Val Thr Leu Tyr Gly Asn Thr Ser Asn Ala Ser Thr Lys Asp Thr
             20                  25                  30

Leu Thr Gly Ser Gly Arg Trp Glu Thr Ala Ile Lys Ile Ser Gln Ala
         35                  40                  45

Gly Trp Thr Lys Ser Glu Ser Ala Val Leu Val Asn Asp Asn Ser Ile
     50                  55                  60

Ala Asp Ala Leu Ser Ala Thr Pro Phe Ala Lys Ala Lys Asp Ala Pro
 65                  70                  75                  80

Ile Leu Leu Thr Gln Ser Asn Lys Leu Asp Ser Arg Thr Lys Ala Glu
                 85                  90                  95

Leu Lys Arg Leu Gly Val Lys Asn Val Tyr Leu Ile Gly Gly Ser Ile
            100                 105                 110

Ala Leu Ser Ser Glu Ile Glu Lys Gln Leu Asn Ala Glu Asn Ile Asn
        115                 120                 125

Phe Glu Arg Ile Ser Gly Asn Ser Arg Tyr Asp Thr Ser Leu Lys Leu
    130                 135                 140

Ala Glu Lys Leu Asp Arg Glu Lys Ser Ile Ser Lys Ile Val Val Val
145                 150                 155                 160

Asn Gly Glu Lys Gly Leu Ala Asp Ala Val Ser Val Gly Ala Ile Ala
                165                 170                 175

Ala Gln Glu Asn Met Pro Ile Ile Leu Ser Asp Ser Glu Asn Gly Thr
            180                 185                 190

Glu Val Ala Asp Asn Phe Ile Asp Ser Lys Asp Ile Ala Lys Ser Tyr
        195                 200                 205

Val Ile Gly Gly Thr Tyr Ser Ile Ser Asn Ser Val Glu Arg Ser Leu
    210                 215                 220

Pro Asn Ala Thr Arg Ile Ala Gly Ser Ser Arg Ser Glu Thr Asn Ala
225                 230                 235                 240

Lys Ile Ile Glu Glu Phe Tyr Lys Asp Thr Asp Ile Lys Asn Ile Tyr
                245                 250                 255

Val Thr Lys Asp Gly Thr Lys Asn Lys Asn Asp Leu Ile Asp Ser Leu
            260                 265                 270

Ala Val Gly Val Leu Ala Ala Lys Asn Ser Ser Pro Ile Val Leu Ala
        275                 280                 285

Gly Asn Lys Leu Asp Thr Thr Gln Lys Asp Val Leu Asn Thr Lys Ile
    290                 295                 300
```

```
Ile Asp Lys Val Thr Gln Ile Gly Gly Leu Gly Asn Glu Asn Val Val
305                 310                 315                 320

Glu Asp Ile Leu Asp Ile Gln Glu Glu Thr Lys Tyr Thr Val Glu Thr
                325                 330                 335

Ile Asp Glu Leu Asn Ala Ala Ile Lys Arg Ala Asp Ala Asn Asp Ile
            340                 345                 350

Ile Lys Phe Lys Pro Glu Lys Glu Lys Thr Ile Asn Asn Ser Phe Ser
        355                 360                 365

Ile Glu Thr Lys Lys Thr Val Thr Ile Glu Leu Asp Gly Arg Tyr Arg
    370                 375                 380

Gln Thr Ile Thr Leu Asp Ile Pro Asn Gly Lys Phe Asn Asn Tyr Ala
385                 390                 395                 400

Glu Ile Glu Gly Gly Val Lys Leu Lys Asn Ile Lys Asn Glu Ser Leu
                405                 410                 415

Val Asn Lys Gly Ser Ile Gln Asp Leu Asp Ile Tyr Asp Glu Asn Gly
            420                 425                 430

Cys Lys Ile Glu Asn Glu Ser Ser Gly Glu Ile Trp Phe Val Thr Ile
        435                 440                 445

Val Glu Glu Ala Asn Asp Val Tyr Ile Val Asn Ser Gly Asp Ile Thr
    450                 455                 460

Lys Ile Ser Asn Asn Ser Ser Ser Thr Ile Ile Arg Asn Ser Gly Asn
465                 470                 475                 480

Ile Asp Thr Val Thr Gly Lys Lys Glu Pro Ala Ile Ser Gly Asn Lys
                485                 490                 495

Pro Lys Val Asn Asp Thr Glu Lys Glu Thr Lys Ala Ala Arg Gly Leu
            500                 505                 510

Asn Pro Arg Val Glu Ala Cys Ser Val Pro Lys Lys Asp Tyr Val Met
        515                 520                 525

Ile Thr Ile Pro Asn Ser Pro Lys Asp Ser Arg Tyr Lys Ile Tyr Tyr
    530                 535                 540

Arg Val Val Tyr Asn Lys Pro Tyr Ala Met Asp Val Gly Asp Lys Ile
545                 550                 555                 560

Asn Ile Gly Glu Trp Thr Val Ala Pro Thr Asp Glu Glu Pro Phe Leu
                565                 570                 575

Glu Lys Ala Lys Asn Gly Cys Tyr Val Glu Ala Val Glu Val Asn Thr
            580                 585                 590

Ser Thr Lys Glu Val Ser Arg Trp Gly Arg Thr Asn Ala Thr Asp Asp
        595                 600                 605

Gly Phe
    610
```

<210> 25
<211> 803
<212> PRT
<213> Clostridium difficile

<400> 25

```
Met Arg Lys Tyr Lys Ser Lys Lys Leu Ser Lys Leu Leu Ala Leu Ser
  1               5                  10                  15

Thr Val Cys Phe Leu Ile Val Ser Thr Ile Pro Val Ser Ala Glu Asn
             20                  25                  30

His Lys Thr Leu Asp Gly Val Glu Thr Ala Glu Tyr Ser Glu Ser Tyr
         35                  40                  45

Leu Gln Tyr Leu Glu Asp Val Lys Asn Gly Asp Thr Ala Lys Tyr Asn
     50                  55                  60

Gly Val Ile Pro Phe Pro His Glu Met Glu Gly Thr Thr Leu Arg Asn
 65                  70                  75                  80

Lys Gly Arg Ser Ser Leu Pro Ser Ala Tyr Lys Ser Ser Val Ala Tyr
                 85                  90                  95

Asn Pro Met Asp Leu Gly Leu Thr Thr Pro Ala Lys Asn Gln Gly Ser
            100                 105                 110

Leu Asn Thr Cys Trp Ser Phe Ser Gly Met Ser Thr Leu Glu Ala Tyr
        115                 120                 125

Leu Lys Leu Lys Gly Tyr Gly Thr Tyr Asp Leu Ser Glu Glu His Leu
    130                 135                 140

Arg Trp Trp Ala Thr Gly Gly Lys Tyr Gly Trp Asn Leu Asp Asp Met
145                 150                 155                 160

Ser Gly Ser Ser Asn Val Thr Ala Ile Gly Tyr Leu Thr Ala Trp Ala
                165                 170                 175

Gly Pro Lys Leu Glu Lys Asp Ile Pro Tyr Asn Leu Lys Ser Glu Ala
            180                 185                 190

Gln Gly Ala Thr Lys Pro Ser Asn Met Asp Thr Ala Pro Thr Gln Phe
        195                 200                 205

Asn Val Thr Asp Val Val Arg Leu Asn Lys Asp Lys Glu Thr Val Lys
    210                 215                 220

Asn Ala Ile Met Gln Tyr Gly Ser Val Thr Ser Gly Tyr Ala His Tyr
225                 230                 235                 240

Ser Thr Tyr Phe Asn Lys Asp Glu Thr Ala Tyr Asn Cys Thr Asn Lys
                245                 250                 255

Arg Ala Pro Leu Asn His Ala Val Ala Ile Val Gly Trp Asp Asp Asn
            260                 265                 270

Tyr Ser Lys Asp Asn Phe Ala Ser Asp Val Lys Pro Glu Ser Asn Gly
        275                 280                 285
```

```
Ala Trp Leu Val Lys Ser Ser Trp Gly Glu Phe Asn Ser Met Lys Gly
    290                 295                 300

Phe Phe Trp Ile Ser Tyr Glu Asp Lys Thr Leu Leu Thr Asp Thr Asp
305                 310                 315                 320

Asn Tyr Ala Met Lys Ser Val Ser Lys Pro Asp Ser Asp Lys Lys Met
                325                 330                 335

Tyr Gln Leu Glu Tyr Ala Gly Leu Ser Lys Ile Met Ser Asn Lys Val
            340                 345                 350

Thr Ala Ala Asn Val Phe Asp Phe Ser Arg Asp Ser Glu Lys Leu Asp
        355                 360                 365

Ser Val Met Phe Glu Thr Asp Ser Val Gly Ala Lys Tyr Glu Val Tyr
    370                 375                 380

Tyr Ala Pro Val Val Asn Gly Val Pro Gln Asn Asn Ser Met Thr Lys
385                 390                 395                 400

Leu Ala Ser Gly Thr Val Ser Tyr Ser Gly Tyr Ile Asn Val Pro Thr
                405                 410                 415

Asn Ser Tyr Ser Leu Pro Lys Gly Lys Gly Ala Ile Val Val Val Ile
            420                 425                 430

Asp Asn Thr Ala Asn Pro Asn Arg Glu Lys Ser Thr Leu Ala Tyr Glu
        435                 440                 445

Thr Asn Ile Asp Ala Tyr Tyr Leu Tyr Glu Ala Lys Ala Asn Leu Gly
    450                 455                 460

Glu Ser Tyr Ile Leu Gln Asn Asn Lys Phe Glu Asp Ile Asn Thr Tyr
465                 470                 475                 480

Ser Glu Phe Ser Pro Cys Asn Phe Val Ile Lys Ala Ile Thr Lys Thr
                485                 490                 495

Ser Ser Gly Gln Ala Thr Ser Gly Glu Ser Leu Thr Gly Ala Asp Arg
            500                 505                 510

Tyr Glu Thr Ala Val Lys Val Ser Gln Lys Gly Trp Thr Ser Ser Gln
        515                 520                 525

Asn Ala Val Leu Val Asn Gly Asp Ala Ile Val Asp Ala Leu Thr Ala
    530                 535                 540

Thr Pro Phe Thr Ala Ala Ile Asp Ser Pro Ile Leu Leu Thr Gly Lys
545                 550                 555                 560

Asp Asn Leu Asp Ser Lys Thr Lys Ala Glu Leu Gln Arg Leu Gly Thr
                565                 570                 575

Lys Lys Val Tyr Leu Ile Gly Gly Glu Asn Ser Leu Ser Lys Asn Val
            580                 585                 590

Gln Thr Gln Leu Ser Asn Met Gly Ile Ser Val Glu Arg Ile Ser Gly
        595                 600                 605
```

```
Ser Asp Arg Tyr Lys Thr Ser Ile Ser Leu Ala Gln Lys Leu Asn Ser
    610                 615                 620

Ile Lys Ser Val Ser Gln Val Ala Val Ala Asn Gly Val Asn Gly Leu
625                 630                 635                 640

Ala Asp Ala Ile Ser Val Gly Ala Ala Ala Ala Asp Asn Asn Met Pro
                645                 650                 655

Ile Ile Leu Thr Asn Glu Lys Ser Glu Leu Gln Gly Ala Asp Glu Phe
            660                 665                 670

Leu Asn Ser Ser Lys Ile Thr Lys Ser Tyr Ile Ile Gly Gly Thr Ala
        675                 680                 685

Thr Leu Ser Ser Asn Leu Glu Ser Lys Leu Ser Asn Pro Thr Arg Leu
    690                 695                 700

Ala Gly Ser Asn Arg Asn Glu Thr Asn Ala Lys Ile Ile Asp Lys Phe
705                 710                 715                 720

Tyr Pro Ser Ser Asp Leu Lys Tyr Ala Phe Val Val Lys Asp Gly Ser
                725                 730                 735

Lys Ser Gln Gly Asp Leu Ile Asp Gly Leu Ala Val Gly Ala Leu Gly
            740                 745                 750

Ala Lys Thr Asp Ser Pro Val Val Leu Val Gly Asn Lys Leu Asp Glu
        755                 760                 765

Ser Gln Lys Asn Val Leu Lys Ser Lys Lys Ile Glu Thr Pro Ile Arg
    770                 775                 780

Val Gly Gly Asn Gly Asn Glu Ser Ala Phe Asn Glu Leu Asn Thr Leu
785                 790                 795                 800

Leu Gly Lys
```

<210> 26
<211> 525
<212> PRT
<213> Clostridium difficile

<400> 26

```
Met Lys Ala Pro Lys Thr Ile Leu Thr Ile Leu Thr Ile Ala Leu Thr
  1               5                  10                  15

Leu Ser Ser Ile Ser Ile Ile Pro Ser Tyr Ala Leu Thr Glu Glu Lys
            20                  25                  30

Leu Ile Gly Asn Gly Arg Tyr Glu Thr Ala Val Lys Ile Ser Gln Lys
        35                  40                  45

Ala Tyr Ser Ser Ser Pro Asn Val Val Leu Val Asn Asp Asn Ser Leu
    50                  55                  60

Ala Asp Ala Leu Ser Ala Thr Pro Phe Ala Lys Ala Lys Gly Ala Pro
65                  70                  75                  80
```

```
Ile Leu Leu Thr Glu Ser Asp Lys Leu Asp Asp Arg Thr Glu Lys Glu
                85                  90                  95

Ile Lys Arg Leu Gly Ala Lys Asp Ile Tyr Leu Ile Gly Gly Thr Ala
            100                 105                 110

Val Leu Asn Lys Asp Ile Glu Asn Lys Leu Lys Gly Asn Gly Leu Asn
        115                 120                 125

Val Glu Arg Ile Asn Gly Lys Asn Arg Tyr Glu Thr Ser Leu Ile Leu
    130                 135                 140

Ala Asn Lys Leu Lys Asp Ile Lys Asp Ile Lys Glu Val Ala Val Val
145                 150                 155                 160

Asn Gly Glu Lys Gly Leu Ser Asp Ala Val Ser Val Gly Ala Pro Ala
                165                 170                 175

Ala Gln Asn Lys Met Pro Ile Ile Leu Ser Asn Pro Lys Asp Gly Val
            180                 185                 190

Glu Ala Phe Asp Lys Phe Ile Arg Asp Glu Lys Val Ile Lys Ala Tyr
        195                 200                 205

Val Ile Gly Gly Thr Asn Ser Val Ser Arg Ala Val Glu Lys Ser Leu
    210                 215                 220

Pro Asn Ala Glu Arg Val Ser Gly Lys Asp Arg Asn Glu Thr Asn Ala
225                 230                 235                 240

Lys Val Ile Glu Lys Phe Tyr Thr Asp Thr Asn Leu Ser Asn Leu Tyr
                245                 250                 255

Val Thr Lys Asp Gly Ser Lys Asn Glu Asn Gln Leu Ile Asp Ser Leu
            260                 265                 270

Ala Val Gly Val Leu Ala Ala Lys Asn Glu Ser Pro Ile Val Leu Val
        275                 280                 285

Gly Asn Lys Leu Asn Thr Lys Gln Arg Asp Ile Leu Ser Thr Lys Lys
    290                 295                 300

Leu Asn Thr Ile Thr Gln Val Gly Gly Asn Gly Asn Glu Glu Ala Phe
305                 310                 315                 320

Asp Glu Ile Lys Ser Leu Gln Glu Lys Thr Val Phe Glu Ala Lys Thr
                325                 330                 335

Val Glu Glu Leu Thr Asp Met Ile Asn Ile Ala Ser Pro Asn Asp Ile
            340                 345                 350

Ile Asn Phe Lys Pro Lys Glu Asn Thr Val Asn Glu Ala Phe Arg Met
        355                 360                 365

Val Thr Asn Lys Pro Ile Thr Val Asn Ile Lys Gly Asp Cys Ser Lys
    370                 375                 380

Thr Leu Thr Val Asp Met Pro Asn Gly Glu Val Asn Asn Tyr Ala Thr
385                 390                 395                 400
```

```
Leu Val Asn Val Ile Val Arg Asn Ile Gly Glu Gly Gly Phe Asn Asn
                405                 410                 415

His Asp Thr Ile Thr Ile Leu Ser Val Arg Asp Lys Asn Gly Arg Val
            420                 425                 430

Ile Glu Asn Thr Arg Asn Ser Asp Ile Asp Thr Leu Met Ile Leu Ala
        435                 440                 445

Ser Ala Asn Asp Thr Lys Leu Ile Asn Asp Gly Tyr Ile Gly Lys Leu
    450                 455                 460

Ile Asp Asn Ser Ser Asn Ser Asp Ile Thr Asn Asn Gly Thr Ile Asp
465                 470                 475                 480

Lys Lys Val Asn Gln Val Glu Asp Leu Glu Ala Lys Val Asp Ser Ile
                485                 490                 495

Glu Lys Ala Ile Asp Ser Ile Ser Gln Lys Val Asn Lys Ile Gln Asp
            500                 505                 510

Ile Leu Asp Lys Leu Gly Phe Leu Lys Lys Phe Leu Ser
        515                 520                 525
```

<210> 27
<211> 679
<212> PRT
<213> Clostridium difficile

<400> 27

```
Met Arg Gly Asp Met Met Lys Lys Thr Thr Lys Leu Leu Ala Thr Gly
  1                 5                  10                  15

Met Leu Ser Val Ala Met Val Ala Pro Asn Val Ala Leu Ala Ala Glu
             20                  25                  30

Asn Thr Thr Ala Asn Thr Glu Ser Asn Ser Asp Ile Asn Ile Asn Leu
         35                  40                  45

Gln Arg Lys Ser Val Val Leu Gly Ser Lys Ser Asn Ala Ser Val Lys
     50                  55                  60

Phe Lys Glu Lys Leu Asn Ala Asp Ser Ile Thr Leu Asn Phe Met Cys
 65                  70                  75                  80

Tyr Asp Met Pro Leu Glu Ala Thr Leu Asn Tyr Asn Glu Lys Thr Asp
                 85                  90                  95

Ser Tyr Glu Gly Val Ile Asn Tyr Asn Lys Asp Pro Glu Tyr Leu Asn
            100                 105                 110

Val Trp Glu Leu Gln Ser Ile Lys Ile Asn Gly Lys Asp Glu Gln Lys
        115                 120                 125

Val Leu Asn Lys Glu Asp Leu Glu Ser Met Gly Leu Asn Leu Lys Asp
    130                 135                 140

Tyr Asp Val Thr Gln Glu Phe Ile Ile Ser Asp Ala Asn Ser Thr Lys
145                 150                 155                 160
```

```
Ala Val Asn Glu Tyr Met Arg Lys Thr Ser Ala Pro Val Lys Lys Leu
                165                 170                 175

Ala Gly Ala Thr Arg Phe Glu Thr Ala Val Glu Ile Ser Lys Gln Gly
            180                 185                 190

Trp Lys Asp Gly Ser Ser Lys Val Val Ile Val Asn Gly Glu Leu Ala
        195                 200                 205

Ala Asp Gly Ile Thr Ala Thr Pro Leu Ala Ser Thr Tyr Asp Ala Pro
    210                 215                 220

Ile Leu Leu Ala Asn Lys Asp Asp Ile Pro Glu Ser Thr Lys Ala Glu
225                 230                 235                 240

Leu Lys Arg Leu Asn Pro Ser Asp Val Ile Ile Ile Gly Asp Asp Gly
                245                 250                 255

Ser Val Ser Gln Lys Ala Val Ser Gln Ile Lys Ser Ala Val Asn Val
            260                 265                 270

Asn Val Thr Arg Ile Gly Gly Val Asp Arg His Glu Thr Ser Leu Leu
        275                 280                 285

Ile Ala Lys Glu Ile Asp Lys Tyr His Asp Val Asn Lys Ile Tyr Ile
    290                 295                 300

Ala Asn Gly Tyr Ala Gly Glu Tyr Asp Ala Leu Asn Ile Ser Ser Lys
305                 310                 315                 320

Ala Gly Glu Asp Gln Gln Pro Ile Ile Leu Ala Asn Lys Asp Ser Val
                325                 330                 335

Pro Gln Gly Thr Tyr Asn Trp Leu Ser Ser Gln Gly Leu Glu Glu Ala
            340                 345                 350

Tyr Tyr Ile Gly Gly Ser Gln Ser Leu Ser Ser Lys Ile Ile Asp Gln
        355                 360                 365

Ile Ser Lys Ile Ala Lys Asn Gly Thr Ser Lys Asn Arg Val Ser Gly
    370                 375                 380

Ala Asp Arg His Glu Thr Asn Ala Asn Val Ile Lys Thr Phe Tyr Pro
385                 390                 395                 400

Asp Lys Glu Leu Ser Ala Met Leu Val Ala Lys Ser Asp Ile Ile Val
                405                 410                 415

Asp Ser Ile Thr Ala Gly Pro Leu Ala Ala Lys Leu Lys Ala Pro Ile
            420                 425                 430

Leu Ile Thr Pro Lys Thr Tyr Val Ser Ala Tyr His Ser Thr Asn Leu
        435                 440                 445

Ser Glu Lys Thr Ala Glu Thr Val Tyr Gln Ile Gly Asp Gly Met Lys
    450                 455                 460

Asp Ser Val Ile Asn Ser Ile Ala Ser Ser Leu Ser Lys His Asn Ala
465                 470                 475                 480
```

```
Pro Thr Glu Pro Asp Asn Ser Gly Ser Ala Ala Gly Lys Thr Val Val
                485                 490                 495

Ile Asp Pro Gly His Gly Gly Ser Asp Ser Gly Ala Thr Ser Gly Leu
            500                 505                 510

Asn Gly Gly Ala Gln Glu Lys Lys Tyr Thr Leu Asn Thr Ala Leu Ala
        515                 520                 525

Thr Thr Glu Tyr Leu Arg Ser Lys Gly Ile Asn Val Val Met Thr Arg
    530                 535                 540

Asp Thr Asp Lys Thr Met Ala Leu Gly Glu Arg Thr Ala Leu Ser Asn
545                 550                 555                 560

Thr Ile Lys Pro Asp Leu Phe Thr Ser Ile His Tyr Asn Ala Ser Asn
                565                 570                 575

Gly Ser Gly Asn Gly Val Glu Ile Tyr Tyr Lys Val Lys Asp Lys Asn
            580                 585                 590

Gly Gly Thr Thr Lys Thr Ala Ala Ser Asn Ile Leu Lys Arg Ile Leu
        595                 600                 605

Glu Lys Phe Asn Met Lys Asn Arg Gly Ile Lys Thr Arg Thr Leu Asp
    610                 615                 620

Asn Gly Lys Asp Tyr Leu Tyr Val Leu Arg Asn Asn Asn Tyr Pro Ala
625                 630                 635                 640

Ile Leu Val Glu Cys Ala Phe Ile Asp Asn Lys Ser Asp Met Asp Lys
                645                 650                 655

Leu Asn Thr Ala Glu Lys Val Lys Thr Met Gly Thr Gln Ile Gly Ile
            660                 665                 670

Gly Ile Glu Asp Thr Val Lys
        675
```

<210> 28
<211> 351
<212> PRT
<213> Clostridium difficile

<400> 28

```
Met Ile Lys Lys Ile Ser Thr Ile Leu Ser Leu Val Leu Leu Ile Ser
  1               5                  10                  15

Ile Ser Ser Thr Ile Gly Val Phe Ala Asp Ala Asn Pro Lys Arg Glu
             20                  25                  30

Leu Ile Glu Gly Ser Ile Pro Glu Ile Ser Thr Glu Leu Asn Lys Arg
         35                  40                  45

Ala Phe Lys Asp Ser Lys Glu Val Ile Leu Val Asn Glu Glu Ser Ile
     50                  55                  60

Val Asp Ser Ile Ser Ala Thr Pro Leu Ala Tyr Ser Lys Asn Ala Pro
 65                  70                  75                  80
```

```
Ile Val Val Thr Lys Ser Lys Asn Leu Gly Arg Val Thr Arg Asn Tyr
                 85                  90                  95

Leu Lys Glu Leu Gly Pro Glu Lys Val Thr Ile Val Gly Gly Leu Lys
            100                 105                 110

Ala Val Ser Lys Asp Ala Glu Arg Asn Ile Glu Lys Met Gly Met Lys
        115                 120                 125

Val Glu Arg Ile Arg Gly Lys Asp Arg Tyr Asp Thr Ser Leu Lys Ile
    130                 135                 140

Ala Arg Glu Met Tyr Arg Thr Val Gly Phe Asp Glu Ala Phe Leu Leu
145                 150                 155                 160

Ser Ser Thr Thr Gly Leu Glu Asn Ala Ile Ser Val Tyr Ser Tyr Ala
                165                 170                 175

Ala Lys Ser Gly Met Pro Ile Ile Trp Ala Lys Asp Glu Gly Phe Glu
            180                 185                 190

Glu Gln Ile Asp Phe Leu Lys Gly Lys Asn Leu Lys Lys Ile Tyr Ala
        195                 200                 205

Leu Gly Asp Ser Lys Glu Phe Ile Ala Glu Ile Asp Ser Asn Leu Lys
    210                 215                 220

Asn Ile Glu Gly Ile Lys Gln Ile Asn Lys Ser Ser Thr Asn Val Asp
225                 230                 235                 240

Leu Ile Lys Lys Phe Tyr Asp Glu Lys Asp Ile Lys Lys Ile Tyr Thr
                245                 250                 255

Ala Arg Leu Asp Phe Gly Ser Arg Ser Asp Val Asn Glu Tyr Ile Ser
            260                 265                 270

Leu Gly Val Val Ser Ala Lys Glu Asn Met Pro Ile Leu Ile Cys Ser
        275                 280                 285

Asp Asn Leu Ser Arg Ala Gln Asp Lys Phe Leu Lys Asp Ser Asn Ile
    290                 295                 300

Asn Asp Val Val Glu Val Gly Tyr Thr Val Gly Asp Tyr Ser Leu Phe
305                 310                 315                 320

Lys Ser Ile Phe Asn Leu Thr Phe Leu Ser Cys Ile Val Leu Ile Leu
                325                 330                 335

Leu Leu Leu Leu Ile Thr Phe Arg Ala Leu Arg Tyr Glu Ser Lys
            340                 345                 350


<210> 29
<211> 631
<212> PRT
<213> Clostridium difficile

<400> 29
Met Leu Ser Asn Lys Lys Arg Ser Met Ala Ile Val Met Ala Gly Ala
  1                 5                  10                  15
```

```
Thr Val Met Ser Ala Ala Ala Pro Ile Phe Ala Asp Asn Thr Val Thr
            20                  25                  30

Glu Asn Val Asp Lys Asn Tyr Thr Val Ser Ala Lys Asp Ser Ala Lys
        35                  40                  45

Leu Ile Glu Glu Val Arg Lys Ala Leu Glu Val Lys Phe Glu Asp Thr
    50                  55                  60

Lys Ala Gly Ala Asn Val Asn Asp Arg Val Tyr Asp Ile Lys Val Asp
65                  70                  75                  80

Asn Val Asn Leu Thr Asn Ala Thr Gln Leu Gln Asn Lys Ile Asn Ser
                85                  90                  95

Leu Thr Glu Gly Gln Ser Leu Lys Val Thr Ile Gln Asp Lys Gly His
            100                 105                 110

Gln Val Leu Gly Gly Lys Val Val Asp Tyr Lys Ile Glu Asn Tyr Lys
        115                 120                 125

Thr Ala Gln Glu Ile Val Asp Ala Val Asn Ala Tyr Asn Ala Thr Leu
    130                 135                 140

Ala Glu Asp Ser Asp Asn Lys Leu Thr Ala Thr Ile Lys Ser Thr Asn
145                 150                 155                 160

Thr Val Glu Val Lys Arg Ala Lys Asp Ser Ala Asn Val Ile Thr Leu
                165                 170                 175

Asn Val Gly Asp Gln His Leu Asp Phe Ser Lys Val Ile Thr Ser Glu
            180                 185                 190

Glu Gly Thr Phe Glu Gly Tyr Glu Lys Arg Tyr Ser Asp Ile Asp Ser
        195                 200                 205

Lys Glu Leu His Thr Val Thr Val Lys Asn Ala Asp Leu Gln Asp Ile
    210                 215                 220

Ser Ala Glu Glu Leu Phe Asp Gly Ile Arg Leu Thr Thr Leu Gly Arg
225                 230                 235                 240

Glu Ile Val Asn Lys Val Lys Asn Gly Tyr Ala Leu Thr Phe Glu Asn
                245                 250                 255

Glu Ala Ile Leu Thr Gln Glu Gln Glu Asp Ser Asp Asp Lys Asp Lys
            260                 265                 270

Pro Glu Lys Ser Ser Phe Asp Ile Val Leu Ser Lys Ala Asn Glu Lys
        275                 280                 285

Pro Glu Thr Ile Ser Val Ser Ser Lys Asn His Lys Leu Val Arg Asp
    290                 295                 300

Leu His Lys Val Leu Thr Asp Val Lys Asp Gly Lys Glu Leu Lys Val
305                 310                 315                 320

Glu Val Leu Ser Gly Asp Ser Arg Phe Thr Thr Ala Val Glu Val Ser
                325                 330                 335
```

```
Lys Glu Arg Phe Lys Asp Gly Glu Ala Glu Ala Ile Ile Leu Val Gly
            340                 345                 350

Glu Asp Ala Ile Val Asp Gly Leu Ala Ser Ala Pro Leu Ala Ser Gln
        355                 360                 365

Lys Asn Ala Pro Ile Leu Leu Ser Lys Lys Asp Ser Leu Pro Ser Glu
    370                 375                 380

Ile Glu Ala Glu Ile Leu Arg Val Leu Gly Ser Asn Leu Ser Ser Lys
385                 390                 395                 400

Lys Ile Tyr Ile Val Gly Gly Glu Ser Lys Val Ser Lys Glu Thr Glu
                405                 410                 415

Glu Lys Leu Ser Lys Leu Gly Val Ser Lys Val Glu Arg Val Ser Gly
            420                 425                 430

Glu Asp Arg Phe Glu Thr Ser Leu Glu Ile Ala Lys Gln Leu Lys Asp
        435                 440                 445

Thr Phe Lys Thr Ala Phe Val Val Gly Gly Asn Gly Glu Ala Asp Ala
    450                 455                 460

Met Ser Ile Ser Ala Arg Ala Ala Gln Phe Gly Ala Pro Ile Ile Val
465                 470                 475                 480

Thr Gly Asn Glu Leu Asp Ala Asn Ala Glu Lys Leu Leu Lys Gly Lys
                485                 490                 495

Glu Leu Glu Ile Val Gly Gly Glu Asn Ser Val Ser Lys Glu Val Glu
            500                 505                 510

Asp Lys Leu Val Asp Ile Asp Leu Asn Asn Lys Val Glu Arg Leu Ala
        515                 520                 525

Gly Glu Asn Arg Lys Asp Thr Asn Ala Lys Val Ile Asn Lys Tyr Tyr
    530                 535                 540

Ala Gly Ala Thr Lys Ala Tyr Val Ala Lys Asp Gly Tyr Val Gly Gly
545                 550                 555                 560

Asn Gly Gln Leu Val Asp Ala Leu Thr Ala Ala Pro Leu Ala Ala Ser
                565                 570                 575

Ser Lys Ala Pro Ile Val Leu Thr Thr Glu Glu Leu Ser Lys Ser Gln
            580                 585                 590

Glu Glu Val Val Glu Leu Arg Leu Lys Asn Ala Thr Lys Leu Val Gln
        595                 600                 605

Ile Gly Glu Gly Ile Ala Lys Asn Ala Ile Glu Lys Ile Ala Glu Lys
    610                 615                 620

Ile Asn Leu Phe Thr Lys Asn
625                 630
```

<210> 30
<211> 477
<212> PRT

<213> Clostridium difficile

<400> 30

```
Met Lys Ser Thr Leu Gly Val Glu Asn Asn Met Lys Asn Ser Lys Lys
  1               5                  10                  15

Ile Leu Ala Ile Gly Leu Thr Leu Phe Leu Val Met Val Asn Thr Pro
            20                  25                  30

Met Val Ser Ala Leu Thr Ser Val Glu Gln Ile Lys Gly Asn Asp Arg
        35                  40                  45

Tyr Glu Thr Ala Ala Lys Ile Ala Asp Lys Gln Asn Tyr Asn Thr Ala
    50                  55                  60

Ile Leu Ile Asn Ser Asp Asn Ser Leu Ala Asp Gly Leu Ser Ala Ser
65                  70                  75                  80

Gly Leu Ala Gly Ala Leu Asn Ala Pro Ile Leu Met Thr Lys Gln Asn
                85                  90                  95

Gln Ile Pro Asn Thr Thr Met Glu Arg Leu Asn Lys Ala Lys Thr Val
            100                 105                 110

Tyr Ile Ile Gly Ser Glu Ser Thr Ile Ser Lys Asn Val Glu Asn Gln
        115                 120                 125

Leu Leu Ser Lys Lys Lys Val Val Gln Arg Ile Phe Gly Glu Asn Arg
    130                 135                 140

Phe Asp Thr Ser Ile Lys Ile Ala Glu Lys Ile Lys Glu Ile Lys Pro
145                 150                 155                 160

Ile Asp Lys Val Ile Ile Ala Asn Gly Phe Thr Gly Glu Ala Asp Ala
                165                 170                 175

Ile Ser Ala Ser Pro Val Ala Ala Arg Asp Gly Val Pro Ile Ile Leu
            180                 185                 190

Thr Asp Gly Asn Ser Val Gly Phe Asp Thr Thr Gly Leu Lys Ser Tyr
        195                 200                 205

Ala Leu Gly Ser Ser Glu Ile Ile Ser Asp Glu Leu Val Lys Ser Thr
    210                 215                 220

Asn Ser Ile Arg Leu Gly Gly Thr Asp Arg Phe Glu Thr Asn Lys Ile
225                 230                 235                 240

Val Ile Gln Glu Phe Tyr Lys Asn Ser Lys Glu Phe Tyr Leu Ser Lys
                245                 250                 255

Gly Leu Gln Leu Thr Asp Ala Leu Ala Ala Ser Thr Ile Ala Lys Asn
            260                 265                 270

Ala Pro Val Val Leu Val Glu Asn Gly Ser Asn Lys Ser Ile Leu Ser
        275                 280                 285

Gly Ala Asp Lys Leu Thr Val Leu Gly Gly Ile Asn Gln Asn Val Ile
    290                 295                 300
```

```
Lys Gln Cys Ile Asn Gln Ala Ser Pro Asn Gln Gln Gly Leu Tyr Tyr
305                 310                 315                 320

Asn Pro Asn Asp Arg Ala Phe Lys Glu Arg Ile Lys Gly Lys Val Tyr
                325                 330                 335

Ala Leu Thr Lys Gln Tyr Arg Lys Glu Asn Gly Val Arg Ala Leu Ser
            340                 345                 350

Val Ala Ser Arg Leu Glu Gly Leu Ala Asn Asp Trp Ser Asn Leu Met
        355                 360                 365

Ala Asn Lys Lys Thr Leu Ser His Thr Ile Asn Gly Lys Asn Ser Tyr
    370                 375                 380

Ser Thr Phe Leu Lys Tyr Leu Asp Trp Ser Glu Ile Lys Pro Gly Tyr
385                 390                 395                 400

Ile Ala Val Gln Gly Glu Asn Ile Ile Lys Tyr Lys Ile Pro Asp Lys
                405                 410                 415

Pro Val Tyr Thr Asn Arg Asp Ala Asp Asp Ile Gly Asn Phe Ile Phe
            420                 425                 430

Asn Glu Trp Lys Thr Asn Pro Glu Glu Gly Thr Asn Met Leu His Lys
        435                 440                 445

Gly Tyr Glu Ile Met Gly Phe Gly Ile Ala Ile Thr Gly Asp Lys Asn
    450                 455                 460

Leu Tyr Ala Thr His Glu Phe Tyr Gly Arg Tyr Lys Glu
465                 470                 475


<210> 31
<211> 626
<212> PRT
<213> Clostridium difficile

<400> 31
Met Asn Lys Arg Lys Ser Phe Ile Arg Thr Ile Ala Val Ser Thr Met
  1               5                  10                  15

Ala Val Ala Val Thr Gly Ser Ala Thr Cys Ala Tyr Ala Ala Pro Val
            20                  25                  30

Leu Gln Gly Thr Lys Thr Tyr Glu Lys Val Asn Thr Ile Asp Ile Ser
        35                  40                  45

Val Asp Ser Val Glu Asn Ile Val Tyr Ser Phe Gln Ala Ser Ile Lys
    50                  55                  60

Val Gln Gly Glu Val Glu Val Val Asp Asn Glu Gln Lys Glu Lys Ile
65                  70                  75                  80

Thr Trp Ser Asp Asn Ile Lys Ser Gln Ile Lys Ser Gly Asn Ala Asn
                85                  90                  95

Ala Thr Cys Arg Ala Glu Tyr Asn Lys Ser Ser Asn Thr Thr Thr Leu
            100                 105                 110
```

```
Asp Ile Tyr Val Thr Ser Asn Glu Asp Leu Leu Asp Gly Asn Arg Leu
        115                 120                 125

Asn Ile Gly Arg Ile Ser Val Lys Lys Ser Gly Ser Asn Ser Asn Ala
    130                 135                 140

Asp Tyr Lys Val Leu Gly Lys Gly Thr Ser Asp Lys Pro Ala Leu Lys
145                 150                 155                 160

Ile Val Thr Tyr Asn Asn Lys Thr Val Asp Tyr Glu Asn Ile Ser Ser
                165                 170                 175

Asp Glu Gly Leu Ile Phe Thr Leu Ile Asn Glu Ser Glu Val Lys Pro
            180                 185                 190

Ile Gly Gly Thr Gly Ser Ser Lys Asn Asp Pro Glu Lys Tyr Lys Val
        195                 200                 205

Glu Lys Ser Glu Ala Leu Glu Tyr Leu Leu Asn Asn Ile Arg Ile Asn
    210                 215                 220

Tyr Ser Ile Val Ser Lys Glu Thr Gln Glu Ser Gly Ser Asn Val Ile
225                 230                 235                 240

Leu Lys Leu Gly Leu Ala Gln Lys Thr Thr Lys Gly Arg Lys Ala Thr
                245                 250                 255

Ile Asn Lys Tyr Val Glu Val Thr Leu Pro Lys Ser Leu Glu Tyr Ile
            260                 265                 270

Val Glu Asn Glu Leu Ser Lys Pro Asp Glu Leu Pro Pro Asp Asn Gly
        275                 280                 285

Ser Gly Gly Asn Asn Gly Gly Gly Ser Asn Ser Gly Gly Ser Ser Ser
    290                 295                 300

Gly Gly Ser Ser Gly Gly Gly Asn Ser Ser Asp Ser Thr Ser Asn Val
305                 310                 315                 320

Thr Val Lys Lys Leu Lys Gly Ala Asp Arg Phe Glu Thr Ala Ile Lys
                325                 330                 335

Ile Ser Gln Ser Gly Trp Thr Lys Ser Asp Thr Val Val Ile Val Asn
            340                 345                 350

Gly Glu Asp Lys Ser Met Val Asp Gly Leu Thr Ala Thr Pro Leu Ala
        355                 360                 365

Ser Val Lys Asn Ser Pro Ile Leu Leu Ser Ser Asn Glu Lys Leu Pro
    370                 375                 380

Gln Lys Thr Val Glu Glu Leu Lys Arg Leu Asn Pro Ser Lys Val Val
385                 390                 395                 400

Val Ile Gly Gly Asn Asn Ser Met Pro Asn Ser Val Val Glu Ala Ile
                405                 410                 415

Lys Ala Val Asn Ser Lys Ile Ser Val Gln Arg Ile Gly Gly Asp Thr
            420                 425                 430
```

```
Arg Tyr Gln Thr Ser Ile Asn Ile Ala Lys Glu Ile Asp Arg Thr Asn
        435                 440                 445

Asn Val Ser Lys Leu Tyr Ile Gly Ala Gly Asn Gly Glu Ala Asp Ser
    450                 455                 460

Leu Ser Ile Ala Ser Leu Ala Gly Lys Glu Lys Thr Pro Ile Val Leu
465                 470                 475                 480

Thr Gln Lys Asp Gly Val Asp Asn Glu Ala Glu Gln Phe Ile Lys Ser
                485                 490                 495

Asn Lys Val Ser Asn Ile Tyr Phe Ile Gly Gly Val Glu Lys Ile Ser
            500                 505                 510

Asn Lys Ala Ile Glu Gln Val Gly Lys Ile Ala Asn Lys Asp Ile Ser
        515                 520                 525

Asn Asn Arg Val Ala Gly Gln Thr Arg Gln Glu Thr Asn Ala Lys Val
    530                 535                 540

Ile Asp Lys Phe Tyr Ser Gln Ser Lys Leu Asp Gly Val Val Val Ala
545                 550                 555                 560

Asn Gln Asp Lys Leu Ile Asp Ala Leu Ala Val Gly Pro Leu Ala Ala
                565                 570                 575

Lys Asn Asn Ser Pro Val Ile Leu Ala Thr Asn Thr Leu Asp Lys Ser
            580                 585                 590

Gln Glu Ser Ser Leu Lys Gly Lys Asn Ser Ser Lys Leu Phe Glu Val
        595                 600                 605

Gly Gly Gly Ile Ala Ser Ser Val Ile Asp Lys Ile Lys Ser Leu Ile
    610                 615                 620

Glu Lys
625
```

<210> 32
<211> 550
<212> PRT
<213> Clostridium difficile

```
<400> 32
Met Glu Asn Asn His Asn Ile Asn Ile Lys Tyr Lys Asn His Gln Gly
  1               5                  10                  15

Asp Met Lys Met Asn Lys Lys Ile Leu Ser Leu Gly Leu Ala Val Ser
             20                  25                  30

Leu Ile Leu Val Asn Phe Lys Ser Val Asn Ala Ser Ser Val Val Glu
         35                  40                  45

Lys Ile Tyr Gly Lys Asp Arg Tyr Glu Thr Ala Ala Lys Ile Ala Asp
     50                  55                  60

Lys Gln Thr Tyr Glu Thr Val Ile Leu Val Asn Thr Glu Lys Ser Leu
 65                  70                  75                  80
```

```
Ala Asp Gly Leu Ser Val Ser Gly Leu Ser Gly Ala Thr Lys Ala Pro
                85                  90                  95

Ile Leu Phe Thr Gln Gln Asn Lys Ile Pro Ala Asp Thr Asn Arg Cys
            100                 105                 110

Leu Lys Asn Ile Lys Lys Ala Tyr Ile Ile Gly Thr Glu Asp Thr Ile
        115                 120                 125

Ser Lys Ser Val Glu Lys Glu Leu Asp Ser Lys Asn Ile Glu Val Lys
    130                 135                 140

Arg Ile Gly Gly Glu Asp Arg Leu Lys Thr Ser Tyr Leu Ile Ala Lys
145                 150                 155                 160

Glu Ile Ala Thr Ile Lys Lys Val Asp Lys Val Leu Leu Thr Asn Ala
                165                 170                 175

Tyr Ser Gly Glu Ala Asp Ala Met Ser Val Ser Ser Val Ala Thr Arg
            180                 185                 190

Asp Gly Ala Pro Ile Ile Leu Thr Asp Gly Lys Ser Val Pro Phe Asp
        195                 200                 205

Val Lys Asn Ile Gln Ser Tyr Cys Ile Gly Ser Glu Glu Ile Met Ser
    210                 215                 220

Asn Pro Leu Val Lys Asn Thr Asn Ser Val Arg Ile Glu Gly Thr Asp
225                 230                 235                 240

Arg Phe Glu Thr Asn Lys Asn Val Ile Asp Tyr Phe Phe Asn Ser Ala
                245                 250                 255

Asp Gly Phe Tyr Val Ser Asp Gly Tyr Gln Leu Val Asp Ala Ile Ala
            260                 265                 270

Ala Ala Pro Leu Thr Lys Asn Ser Pro Met Val Leu Val Asn Asp Gly
        275                 280                 285

Ser Asp Lys Thr Val Leu Glu Gly Ala Lys Asn Ile Thr Ser Val Gly
    290                 295                 300

Glu Ile Asn Glu Lys Val Ile Gln Gln Cys Ile Asn Ala Ser Lys Ser
305                 310                 315                 320

Asn Gly Gln Pro Pro Thr Ile Thr Val Gly Ser Thr Glu Val Tyr Lys
                325                 330                 335

Gly Glu Lys Phe Asp Thr Gly Lys Leu Asn Ile Val Ala Lys Asp Asn
            340                 345                 350

Thr Gly Lys Val Leu Pro Ile Glu Val Asp Gly Phe Ile Asp Thr Asn
        355                 360                 365

Arg Val Gly Thr Tyr Ile Leu Thr Leu Lys Ala Thr Asp Glu Trp Gly
    370                 375                 380

Lys Ser Thr Gly Lys Arg Val Glu Ile Lys Val Leu Asp Asp Lys Ser
385                 390                 395                 400
```

His Asp Tyr Asn Ser Pro Glu Phe Lys Lys Met Val Ser Thr Glu Met
405 410 415

Tyr Asn Leu Ile Asn Ser Tyr Arg Lys Glu Lys Gly Lys Glu Pro Leu
420 425 430

Val Val Ser Ser Arg Leu Glu Gly Met Ala Asn Ala Trp Ser Lys Tyr
435 440 445

Met Met Asp Lys Lys Val Phe Ala His Tyr Ile Asp Gly Lys Asn Ala
450 455 460

Pro Gln Val Phe Ser Glu Phe Gly Met Arg Ser Glu Glu Asn Ile Ala
465 470 475 480

Tyr Ile Tyr Ile Asp Ser Lys Asn Val Gln Thr Thr Gln Asp Ala Lys
485 490 495

Asp Leu Ala Lys Ala Ile Phe Glu Val Trp Lys Lys Ser Pro Glu Tyr
500 505 510

Asn Ala Asn Met Leu Ser Asp Glu Phe Tyr Ser Thr Gly Phe Gly Leu
515 520 525

Tyr Ile Leu Ser Asp Gly Gln Val His Ala Thr Gln Glu Phe Leu Asn
530 535 540

Gly Asn Glu Gly Ser Leu
545 550

<210> 33
<211> 528
<212> PRT
<213> Clostridium difficile

<400> 33
Met Lys Val Asn Lys Arg Val Leu Ser Ile Gly Leu Ala Ile Ser Leu
1 5 10 15

Ile Met Ala Gly Ala Pro Asn Ile Asn Ala Leu Ser Ser Ile Glu Lys
20 25 30

Ile Gln Gly Lys Asp Arg Tyr Glu Thr Ala Ala Lys Ile Ala Gln Lys
35 40 45

Gln Thr Tyr Glu Asn Val Val Leu Val Asn Thr Asp Asn Thr Leu Ala
50 55 60

Asp Gly Leu Ser Ala Ser Gly Leu Ala Gly Thr Val Lys Ala Pro Ile
65 70 75 80

Leu Leu Ser Gln Arg Asn Ser Ile Pro Ser Asp Thr Glu Lys Met Leu
85 90 95

Lys Asp Val Lys Lys Val Tyr Ile Ile Gly Thr Glu Asp Ser Ile Gly
100 105 110

Lys Ser Val Glu Asn Glu Leu Lys Gln Lys Gly Ile Glu Val Lys Arg
115 120 125

```
Ile Gly Gly Asn Asp Arg Ile Glu Thr Ser Tyr Leu Ile Ala Lys Glu
    130                 135                 140

Ile Ala Ser Ile Lys Pro Ile Asp Lys Val Phe Ile Thr Asn Gly Tyr
145                 150                 155                 160

Thr Gly Glu Ala Asp Ala Met Ser Ala Ser Ser Val Ala Ser Arg Asp
                165                 170                 175

Gly Ala Pro Ile Ile Leu Thr Asn Gly Lys Asn Val Pro Phe Glu Lys
            180                 185                 190

Lys Glu Gly Val Gln Cys Tyr Ala Leu Gly Ser Glu Glu Ile Ile Ser
        195                 200                 205

Asn Asp Leu Val Lys Lys Thr Asn Ser Val Arg Leu Ala Gly Glu Asp
    210                 215                 220

Arg Phe Glu Thr Asn Lys Lys Val Ile Lys His Phe Tyr Ser Ser Ala
225                 230                 235                 240

Lys Glu Phe Tyr Leu Ser Lys Gly Tyr Gln Leu Val Asp Ala Val Ala
                245                 250                 255

Gly Ser Ser Ile Ala Lys Asn Ala Pro Ile Val Leu Val Asp Gly Asn
            260                 265                 270

Ser Asp Lys Ser Val Leu Arg Ser Ala Asp Lys Ile Thr Ala Leu Gly
        275                 280                 285

Gly Ile Asp Glu Lys Thr Leu Glu Gln Cys Leu Ser Ala Ser Ser Leu
    290                 295                 300

Asp Ala Ser Ala Pro Thr Ile Thr Val Gly Asn Leu Asn Ile Tyr Gln
305                 310                 315                 320

Gly Asp Lys Phe Asp Ile Ser Lys Leu Asn Ile Val Ala Lys Asp Ser
                325                 330                 335

Asn Gly Asn Asp Leu Thr Pro Glu Leu Ile Gly Asn Ile Asn Thr Asp
            340                 345                 350

Lys Val Gly Lys Tyr Lys Val Thr Ile Lys Ala Thr Asp Ile Gly Gly
        355                 360                 365

Lys Thr Thr Ser Ile Ser Val Glu Val Asn Val Leu Glu Tyr Lys Thr
    370                 375                 380

Asn Asp Met Asn Ser Ser Glu Phe Lys Arg Met Val Ser Ser Glu Met
385                 390                 395                 400

Tyr Ser Leu Val Asn Ser Tyr Arg Lys Glu Lys Gly Lys Glu Pro Leu
                405                 410                 415

Gln Val Ser Glu Asn Leu Gln Gly Met Ala Asn Tyr Trp Ser Lys Tyr
            420                 425                 430

Met Ala Asp Lys Gly Glu Phe Ala His Val Ile Asn Gly Lys Asn Ala
        435                 440                 445
```

```
Ala Glu Val Phe Ser Gly Gly Ile Arg Ser Glu Glu Asn Ile Ala Phe
    450                 455                 460

Val Pro Leu Thr Thr Lys Ser Thr Tyr Thr Thr Lys Asp Ala Arg Glu
465                 470                 475                 480

Val Ala Asn Val Ile Phe Thr Val Trp Lys Lys Ser Asp Lys Tyr Asn
                485                 490                 495

Glu Asn Met Leu Ser Ser Lys Phe Ala Tyr Thr Gly Phe Gly Leu Tyr
            500                 505                 510

Ile Leu Pro Asp Gly Gln Val Tyr Ala Thr Gln Glu Phe Leu Asn Lys
        515                 520                 525
```

<210> 34
<211> 781
<212> PRT
<213> Clostridium difficile

<400> 34

```
Met Ser Val Ile Asp Ser Ile Leu Asp Lys Ala Asp Glu Gln Glu Ile
  1               5                  10                  15

Lys Lys Leu Asn Val Ile Val Asp Lys Ile Asp Ala Leu Glu Asp Ser
            20                  25                  30

Met Lys Asn Leu Ser Tyr Glu Glu Leu Lys Asp Met Thr Ala Ile Phe
        35                  40                  45

Lys Asn Arg Leu Lys Lys Gly Glu Thr Leu Asp Asp Ile Leu Pro Glu
    50                  55                  60

Ala Phe Ala Val Val Arg Glu Val Ser Lys Arg Lys Leu Gly Met Arg
65                  70                  75                  80

Gln Tyr Arg Val Gln Leu Ile Gly Gly Ile Val Ile His Gln Gly Lys
                85                  90                  95

Ile Ala Glu Met Lys Thr Gly Glu Gly Lys Thr Leu Val Glu Val Ala
            100                 105                 110

Pro Val Tyr Leu Asn Ala Leu Thr Gly Lys Gly Val His Val Ile Thr
        115                 120                 125

Val Asn Asp Tyr Leu Ala Glu Arg Asp Lys Glu Leu Met Ser Pro Val
    130                 135                 140

Tyr Glu Ser Leu Gly Met Thr Val Gly Val Ile Ile Ser Asn Gln Asp
145                 150                 155                 160

Pro Asn Ile Arg Lys Gln Gln Tyr Lys Cys Asp Ile Thr Tyr Gly Thr
                165                 170                 175

Asn Ser Glu Phe Gly Phe Asp Tyr Leu Arg Asp Asn Met Val Pro Asp
            180                 185                 190

Leu Ser His Lys Val Gln Arg Glu Leu Asn Phe Ala Ile Val Asp Glu
        195                 200                 205
```

```
Val Asp Ser Ile Leu Ile Asp Glu Ala Arg Thr Pro Leu Ile Ile Ala
    210                 215                 220

Gly Asp Gly Asp Glu Asp Leu Lys Leu Tyr Glu Leu Ala Asn Ser Phe
225                 230                 235                 240

Val Lys Thr Val Lys Glu Glu Asp Phe Glu Leu Asp Arg Lys Asp Lys
                245                 250                 255

Thr Ile Ala Leu Thr Ala Ser Gly Ile Ser Lys Ala Glu Ser Phe Phe
            260                 265                 270

Gly Ile Thr Asn Leu Thr Asp Ile Lys Asn Ile Glu Leu Tyr His His
        275                 280                 285

Ile Asn Gln Ala Leu Arg Gly His Lys Leu Met Glu Lys Asp Val Asp
    290                 295                 300

Tyr Val Ile Ser Asn Gly Glu Val Met Ile Val Asp Glu Phe Thr Gly
305                 310                 315                 320

Arg Val Met Asp Gly Arg Arg Tyr Thr Asp Gly Leu His Gln Ala Ile
                325                 330                 335

Glu Ala Lys Glu Gly Val Glu Ile Lys Asn Glu Ser Lys Thr Met Ala
            340                 345                 350

Thr Val Thr Tyr Gln Asn Phe Phe Arg Leu Tyr Glu Lys Leu Ser Gly
        355                 360                 365

Met Thr Gly Thr Ala Lys Thr Glu Glu Gly Glu Phe Glu Ser Ile Tyr
    370                 375                 380

Lys Leu Asn Val Val Gln Ile Pro Thr Asn Arg Pro Val Ile Arg Ala
385                 390                 395                 400

Asp Leu His Asp Lys Val Phe Lys Thr Glu Glu Glu Lys Tyr Ser Ala
                405                 410                 415

Val Val Glu Glu Ile Ile Arg Ile His Lys Thr Arg Gln Pro Ile Leu
            420                 425                 430

Val Gly Thr Val Ser Val Glu Lys Ser Glu Lys Leu Ser Lys Met Leu
        435                 440                 445

Lys Lys Gln Gly Ile Lys His Gln Val Leu Asn Ala Lys Gln His Asp
    450                 455                 460

Lys Glu Ala Glu Ile Ile Ser Lys Ala Gly Lys Leu Asp Ala Ile Thr
465                 470                 475                 480

Ile Ala Thr Asn Met Ala Gly Arg Gly Thr Asp Ile Ser Leu Gly Ala
                485                 490                 495

Gly Asp Lys Glu Glu Glu Gln Glu Val Lys Asp Leu Gly Gly Leu Tyr
            500                 505                 510

Val Ile Gly Thr Glu Arg His Glu Ser Arg Arg Ile Asp Asn Gln Leu
        515                 520                 525
```

```
Arg Gly Arg Ser Gly Arg Gln Gly Asp Pro Gly Thr Ser Arg Phe Phe
    530                 535                 540

Val Ser Leu Glu Asp Asp Val Ile Lys Leu Tyr Gly Gly Lys Thr Ile
545                 550                 555                 560

Glu Lys Leu Met Lys Arg Thr Ser Ser Asn Glu Asn Thr Ala Ile Glu
                565                 570                 575

Ser Lys Ala Leu Thr Arg Ala Ile Glu Arg Ala Gln Lys Gly Val Glu
            580                 585                 590

Gly Lys Asn Phe Glu Ile Arg Lys Asn Val Leu Lys Tyr Asp Asp Thr
        595                 600                 605

Ile Asn Glu Gln Arg Lys Val Ile Tyr Asn Glu Arg Asn Lys Val Leu
    610                 615                 620

Asn Asp Glu Asp Ile Gln Glu Asp Ile Gln Lys Met Val Lys Asp Ile
625                 630                 635                 640

Ile Gln Glu Ala Gly Glu Thr Tyr Leu Ile Gly Arg Lys Arg Asp Tyr
                645                 650                 655

Tyr Gly Tyr Phe Lys His Leu Tyr Ser Thr Phe Met Pro Ala Asp Thr
            660                 665                 670

Leu Leu Ile Pro Gly Val Asp Lys Lys Ser Val Gln Glu Ile Ile Asp
        675                 680                 685

Ser Thr Tyr Glu Ile Ser Lys Arg Val Tyr Asp Leu Lys Lys Met Met
    690                 695                 700

Leu Gly Ile Asp Lys Val Ala Glu Leu Glu Lys Thr Val Leu Leu Lys
705                 710                 715                 720

Val Val Asp Gln Tyr Trp Ile Asp His Ile Asp Ala Met Glu Gln Leu
                725                 730                 735

Lys Gln Tyr Ile Gly Leu Lys Ser Tyr Ala Gln Lys Asp Pro Phe Lys
            740                 745                 750

Glu Tyr Ala Leu Glu Gly Tyr Asp Met Phe Glu Ala Leu Asn Lys Asn
        755                 760                 765

Ile Arg Glu Ala Thr Val Gln Tyr Leu Tyr Lys Phe Asn
    770                 775                 780
```

<210> 35
<211> 48551
<212> DNA
<213> Clostridium difficile

<400> 35

```
aggtatgatt tatgttttaa taggagttat agctatagca ggactatttg ctataaaaaa 60
gtttttagta acaaaagaca agagtgctct taatataatg agtattcatg atgaagaagc 120
aattaaggaa caacagaaca aagaagacga agataaatct atttagacaa taaattatat 180
gaatagctta ttacaaatat atcttaagaa ataatgagaa ataaatatga ttaaggagaa 240
ctttatgtac gcagtaagat acgaattgat aaagacttgt aagcagagtg gtgcaaggtt 300
aggaagatta cacactccac atgggattat tgaaacacca atattcatgc cagttggaac 360
```

tcaagcaact gtaaaatcta tgacaccaga agaattaaaa gagattggtt cacaaataat 420
acttagtaac acataccatt tatatatgag accaggtcat gaacttataa aaagagctgg 480
tggtcttcat aagtttatga attgggataa accaatactt acagatagtg gagggttcca 540
ggtatttagt ctagggcctt taagaaagat aaaagaagaa ggtgtagaat ttaggtcaca 600
cttagatggt tcaaaacatt ttctaactcc agaaaaagct atggaaatac aaaatgcatt 660
gggttcagac ataatgatgg catttgatga gtgtgcacca tacccatcag atagagaata 720
tgtaaaaaat tctttagaaa gaacaacaag atggttaaaa agatgtaaag atgctcataa 780
taataccgac aagcaagctt tatttggtat aatacaaggt gggatgtata aagacttaag 840
ggaacaatca gcaaaagaaa taactaatat tgatttacca ggatatgcta ttggtggact 900
tagtgttgga gaaccaaaac cacttatgta tgacgtgtta gagcatacaa ctccacttat 960
gccaaaggat aaaccaagat acctaatggg agttggtagt ccagatgact tagtagaagg 1020
tgttataaga ggagtagaca tgtttgattg tgttctacct actcgtatcg ctagaaatgg 1080
aactgcaatg actagccaag gtaaagtagt agttagaaat gcaacttatg ctgaagactt 1140
tactccactt gaccctgaat gtgattgtta tgcttgtaaa aactattcaa gagcttatat 1200
aagacacctt ataaaagcaa atgaaatact gggtgcaaga ttaataacta ctcataatct 1260
tcacttttta ttaaatttaa tgaaacaaat aagacaagcg ataatggaag atagattact 1320
tgattttaga aatgaatttt ttgcaaaata tggatatgaa atataggttt tgcactatta 1380
attataatca ttttaatata taatattcaa tagtactcaa agatagagaa agtaaattgt 1440
tatataaaca tttattatga tttacaaaag agtataattt acaaaggagt gaaaaattaa 1500
tgccacagca acaaataatt atgagtatag gtctatgggt agtagtaatc gcaatctttt 1560
atttcctaat gattagacct caaaagaaaa aagacaaaca attaaaagaa atgagaagta 1620
gcctaagtgt aggagacaag gtcattacaa taggtggaat agtagctaat gtagctaaag 1680
ttgaagatga tgttgtaata ttagaattag gaccaaatag aactaaagtt ccatttgaaa 1740
aatgggctat aggaactgtt aaatctaaaa aagaagaaat agaagaagac taaattagta 1800
atttaaaagc cttttctcga atatatatta atttgagagg aggtttttt atgggaaaat 1860
ctaattatat ttttaaaggt ctaggatatg catatataat aactttggct gttttattag 1920
tgtataactt atttttgaca tttacagata ttggtggaga taatataact atggtctcat 1980
catttataac tactatttct gcagctatag gagggtttta cacttcaaaa cacatgaaag 2040
aaaaaggact tatgtatggg cttttagtag gacttttgta catagtatgt atttttttga 2100
ctgtgttttt agcacaagaa aagtttgtat ttgaagtagg gatgatttat aagttattgc 2160
taatatcagc agcaggaggc attggaggcg tattgggtgt caactttaaa tagagtttat 2220
attaaagtag ttttatgata taatcaatag gaaacaatta gaaatggagg cgttgaaaat 2280
ggaaaatata catgtaaaaa ctttatctca agctacttta aagcaaagtg ctgcaaaagg 2340
tggatgtgga gagtgtcaaa cttcttgtca atcagcttgt aaaacttctt gtacagttgc 2400
aaatcaagag tgtgaaagat agaaaattat tgatttagtt gtttgtaaga gggtatctca 2460
aagtgattct tttatcatta tgaggtatcc tcttttattt tcaattatta tataaaaata 2520
tattctattt atgtaaaatt aaatttgatt tcttatcgtt taattatgca aaaatatcta 2580
atcatataag agcattcatt ttaaagtagc tactcttata ttttatttt tatcaattac 2640
atgtattaga ataatttta attttaaact ctaatttatg atgattttta taatatgtta 2700
tatattttaa tatatggaga tagatattaa tttagtgata tatatttaag taataagagt 2760
aaagcaaaaa ttagtgcata gtagtatata aaaatggagt ttgaatactt ttgttagtat 2820
atgattgttt agtagaaaat cactataaaa taggatttct atctttgagc ataattatag 2880
agagctaagt agggggaagg ggttcaagaa gatattatta gtgtttgtgt tagtaaatta 2940
tgtaaacaca taaaatatat tgaaaaataa tattatatat aatattattt atgtgtggga 3000
aatacgaata taaaaaataa aaaataatat tcggaggaca taaatatgtt aagcaacaaa 3060
aagagaagta tggcaatagt aatggcaggt gctacagtta tgagtgcagc agcaccaata 3120
ttcgcagaca atactgtgac agaaaatgtt gataaaaatt acacagtaag tgcgaaagat 3180
tctgctaaac taatcgaaga ggttagaaaa gcattagaag ttaagtttga agatacaaaa 3240
gcaggagcaa atgtaaatga tagagtttac gatataaaag tagacaatgt aaacttaact 3300
aatgcaactc aattacaaaa taaaataaac tctttaacag aagggcaaag tttaaaagtt 3360
actattcaag ataaagggca tcaagtgtta ggtggaaaag tagttgacta taaaattgaa 3420
aactacaaaa ctgctcaaga aatagtagat gcagttaatg cttacaatgc aactttagca 3480
gaagattctg ataacaagtt aacagcaact ataaaatcta caaatacagt tgaagtaaaa 3540
agagcaaaag attcagcaaa tgttattaca ttaaatgtag gagaccaaca tttagatttc 3600
tcaaaagtga taacatcaga agaaggtact tttgaaggtt atgaaaagcg ttatagtgac 3660
atagattcaa aagaattaca tacggtaaca gttaaaaacg cagatttaca agacatatca 3720
gcagaagaat tatttgatgg aattagatta actacattag gaagagaaat agttaataaa 3780
gttaaaaatg gatatgcttt aacatttgaa aatgaagcaa tccttactca agaacaagaa 3840
gatagtgatg ataaagataa accagaaaaa tcaagctttg atatagtttt aagcaaggca 3900
aatgaaaagc cagaaacaat aagtgtttca agtaaaaacc ataagttagt tagagactta 3960
cacaaagtat taactgatgt aaaagatgga aaagaattaa aagtagaagt tttatctggt 4020

gattcaagat ttacaacagc agtagaagtt agtaaagaga gattcaaaga tggagaagca 4080
gaagctataa ttttagttgg agaagatgct atagttgatg gattagcatc agcaccactt 4140
gcctctcaaa aaaatgcacc aatattatta tctaaaaaag attcactacc atcagaaata 4200
gaagctgaaa tattaagagt acttggaagt aatctatctt ctaagaaaat atatatagta 4260
ggtggagaat ctaaagtatc aaaagaaact gaagaaaaac tttctaaatt aggtgtaagt 4320
aaagttgaga gagtttctgg agaagataga tttgaaactt ctttagaaat agcaaaacaa 4380
ttaaaagata catttaagac tgcttttgta gtaggtggaa atggagaagc tgatgctatg 4440
agtatatcag ctagagctgc tcaatttggt gctccaataa tagttacagg taacgaatta 4500
gatgcaaatg ctgaaaaatt attaaaagga aaagaattag aaatagtggg tggagaaaat 4560
tctgtatcaa aagaagttga agacaaatta gttgatatag atttaaataa taaagttgaa 4620
agattagctg gagaaaatag aaaagatact aatgctaaag taatcaataa atactatgca 4680
ggtgcaacta aagcatatgt agcaaaagat ggttatgtag gtggaaatgg acaattagtt 4740
gatgcactta cagcagcacc acttgcagct agttcaaaag ctccaatagt attaactaca 4800
gaagaacttt ctaaatcaca agaagaagta gttgagttaa gacttaagaa tgctactaag 4860
ttagtacaaa taggtgaagg aatagctaaa aacgctattg aaaaaatagc agaaaaaata 4920
aacttattta ctaagaacta ataagatata ggaattaata ttaaaagttt atttagataa 4980
tatatgatta tacaatgcgt taatatgtaa gacatagtgt agatattata aaatatacaa 5040
ataacaagat aaaataataa aaataaggtg tctcatttag aattagattg agatacctta 5100
ttttattttt tgattataat gttgagaaat tcatattgca aaaattatta taaaaagata 5160
cttcttataa tgattaaaca tattactaac ataaattatt ttgaacttat aacttaatta 5220
aataaaataa attataatgt tttgaattta tagagaataa aaactttagg tatattgttt 5280
gagttgtatg gataactatg gtatcatata taaaatatgg taacaaaatt gtaacttttg 5340
attaaaaata agtatgagta ctaataaaaa atatgaaaag tacattgggg gtagaaaata 5400
atatgaagaa tagtaaaaag atattagcta taggacttac actattttta gtaatggtaa 5460
atactcctat ggtaagtgcg ttgacatcag ttgagcaaat aaaagggaat gacagatatg 5520
agacagcagc aaaaatagca gataagcaaa actacaatac agcaatacta atcaactcag 5580
ataatagctt agcagatggt ctaagtgcaa gtggtttagc tggagcctta aatgcgccta 5640
ttttgatgac aaaacaaaat cagattccaa acacaactat ggaaagatta aacaaagcaa 5700
agactgtgta tataataggc tcagaatcaa caataagtaa aaatgttgag aatcagttac 5760
tatccaaaaa gaaggttgta caaagaatct ttggtgaaaa tagatttgat acaagtataa 5820
aaatagctga aaaaattaaa gaaattaagc caatagataa agtaattata gccaatggat 5880
ttacaggaga agcagatgcg ataagtgcat caccagtagc tgctagagat ggagtaccta 5940
taatacttac agatggaaat agtgtgggat ttgacacgac aggcttaaag agttatgcgc 6000
ttggttctag tgaaataata agcgacgaat tagttaagag cacaaattct attagattgg 6060
gtggcacaga tagatttgaa actaataaga tagttataca agagttttat aaaaatagta 6120
aagaatttta tctaagtaaa ggcttacaat taacagacgc tttggctgcc tctacaatag 6180
caaaaaatgc tcctgtagtt ttagttgaaa atggcagtaa taagagtatt ttaagtggtg 6240
cagacaaatt gactgtacta ggaggcataa atcagaatgt tatcaagcaa tgtataaatc 6300
aagcttcacc aaatcaacaa ggtttatatt acaatccaaa tgatagagca tttaaagaga 6360
gaataaaggg taaagtatat gctcttacaa agcagtatag aaaagagaat ggagtaagag 6420
cgttatctgt agcaagcaga ttagagggtc tagcaaatga ttggtctaat ttaatggcaa 6480
ataagaagac tttatctcat acgataaatg gaaaaaattc atattctact tttttgaaat 6540
acttagattg gagtgagatt aaacctggat atatagcagt tcaaggtgaa aatataatta 6600
aatataaaat tccagataaa cctgtatata caaatagaga tgcagatgat ataggaaact 6660
ttatttttaa tgagtggaag acaaatccag aagaaggaac taatatgttg cataaagggt 6720
atgaaataat gggatttggt atagctataa caggagataa gaatctatat gctacacatg 6780
aattttatgg aagatataaa gaataagaaa atatacacaa aattatgctt gtaaggaagc 6840
tatataaata attaaattat aatggaaaat agcatttata ttaataaatt aatataaatg 6900
ttatttttt ataatttaat gtaaaaaaat tgtattaaaa agtttaaata tagtataaaa 6960
tgggtaagaa agtagtgaaa attaaatggc aaaattaggg aaatctaatg acgcaaaact 7020
atagggatta aagcttttaa agccatatca gccagttgct aaaaagagta aagtcttttt 7080
attttataat taagattctc tttgagtgtc ttttttatt gtctgaaact tatagtattt 7140
aaattcatga aaaattagat atttattaag ggaggattgt tatttcatga aaaaggcaat 7200
atcttgtgta ctagcagtat ctatgtgtag tgcaccacta aacgtttttg cagaacctat 7260
attagagggg aagctaaggg cagttgaaaa atcagttaca gaaaagttaa gggggagtct 7320
tgaagtagat ttaaattttt cactgccaat aaaaaatagc gagtcagaaa gtatgactaa 7380
tatgcagtta cgattaaaag atgatagtaa caatagtgga ataataaagc ttggagaaaa 7440
agattcagga acagtaaatg ttgggacaga taatatagaa tatactataa aaaaattatc 7500
tgcaagtaga agtgaaataa aaaataaaga tgaaaatgta tcatattata atattgtatt 7560
taataattta cctgttggaa aatacaatgt agaagtctct ggagctggat ttaagaacaa 7620
agaaattaaa gacattgata tatcaagtta ttcacaaaga gtcttattaa gcaatagagc 7680

aagtgtagat aataaaaata aacctttaaa ttatgaacta ttcttaatgg gtgatgtaaa 7740
tggagacggt aaagtagata agagtgatta taataaggtt ttagagaata tagattcaaa 7800
taaaagagaa tttgatttaa atagagatgg aaaggtagat attgttgact tagattatgt 7860
acagaaaaat ttaggacaaa gtgaagatgc aaaaagccta gaaagtatag tatctaccaa 7920
tcctatagta gatacaagta aggttgaatt aaaaggaaat agtgatgtag atattaatgg 7980
taatgtagaa aatttattca gtggagaaaa taatggagtt acaatatctt caaaaaatga 8040
tgaaatatct gaggaaaaac cagcagtaat ggatatagct ttatcagaag cacttaatat 8100
ggaagtcatc aatataaaag ctccagaagg tactgctcca gcaaaaggaa tagttgaggt 8160
tactgacaaa gatggtaata ttcaacagat tccttatgaa aacaaagaat atatgaaaaa 8220
cttaaaagat gaggagcaac aagaggaagt gaataacgaa caagaagaag taaatcaaga 8280
agaagtaaat aataaagatg aaaatataaa ccagactgat aataactcaa ataataaaga 8340
tgaaaataat gataaacaac aggaaaatag cgataaacaa gaagcaaact ctaataccaa 8400
tcaagataaa aataatgata aagaaaattc atcaatggat aaacctaatt cagatgaaga 8460
tactccaaaa gaaggggata atatcagtaa agatgataac tttacaacat ctacatatag 8520
taggtcagca ttagaaacat atgcagatgt tcaaaatagt gatatagtaa taaatcttgg 8580
taaacaagta gcagtaaaaa agataactat aaaaataaca gctacaacag caaatgatag 8640
aaatcttgca gaaatatcaa aagttgaatt tttaaacaat gtatataaag aaataccaaa 8700
gccagagatg aacataccta aaataaataa aatattgact agcactgctg ttggaaatga 8760
aaaagtatct ttttattgga ataatgaagt aaatgtatca gcatacgaac ttatactaga 8820
caaaattgat agtaagggaa aagtattatc tactaaaaaa ttacaaacca gtaaaaatac 8880
aatagaagtt agagatatag aagcatatgg actatataga gttagcatac aatctttaaa 8940
tggagattgg tctagtggtt ataaagatgc aacaccagaa aacttagacg gagtaccaga 9000
aaatgtagat tcaaactata aacctataca atttagccca gatagtatat tagaatttca 9060
agttgttcca gacacaaaac cagaaccacc agaaggtata aatgttaaag gtaaatttaa 9120
atctcttgat gtaagctgga aaaatcataa acaagctaaa gattttgact tgtactataa 9180
agaaaaaggc tctaatgggt cttggataaa atacaatgat aagcttattc gtgggacaag 9240
ttcaaagata gaaggtctta aggataatac aacatatgaa gttagaatga cagctactaa 9300
ccacttggga actagtggaa tgtctaaaac atatataggt acaactgaaa gcatggatgc 9360
accaacaaca cctaactata gattaattaa tactcctaaa gagaatggtg gatttggaat 9420
accaacaaat cacattgtag atgttactta tccagcagga tatgacaaaa gtgaatatag 9480
tgataaaaat ccatttgata aattcaatgt agtagatggg gactttacaa cacattggac 9540
attcccaaca tataatgcag ctgagggaac gaatagagga ccaataataa cttttgataa 9600
atcttatact atggatactt ttatggctac acctcgtttt gatgtttctg ataagacaat 9660
aaatgacttt gaggtacgat attgggatga aaatggtcaa ttacacaatc tgggggatgt 9720
tccttttgaa acaaagacag accctgtaaa taataataag tatgtaatgg ttagattaaa 9780
agaaccaata actgctacca agatacaagt aagtgtttgt tcatcctatt catataagtc 9840
taatatatca gaacttaagt tttatgagta tgacagtata gaagatgatg taagaaatct 9900
atttatggat gacctacaag ttgaattaaa agaagatgta actcaagaaa aaataactga 9960
attaaaagat agattaaata ctccagataa agctagtgga gaatatcatc catataagac 10020
agttattgag agagaattaa aattagctca agacttatat aatgatagag ccacacttac 10080
tgatgtaact actgtaaatc aagaaataaa taaattagga aaatataaag ataaaaacgg 10140
acagtatact gtaaattcca ataatttagg tatgcaaaat gactggcagg cattaggagt 10200
agctgctaga gcaggagatg aaatcacagt atatgttgga agtaaaagtg gaaaaactcc 10260
taagttgata tatacacagt actatggaga atctggtgca tataagagtg gagaaataaa 10320
tcttaacact ggaaaaaatg taattcaatt aagcaaactt catagtttag atatagaatg 10380
tggaggaagt ttatacataa gatacgctga agatactcca agtggaggag atataaaagt 10440
acgtgtagca ggagctacta agataccaca ccttaatcta aatggaacaa taaatgataa 10500
gagtcctgcg ggagtttctg aatctaagaa aaagataaaa gcatatatag aagaattagc 10560
tagatataaa gatgctgtaa aaggagaacc atattatcca ggaacaaatg ttggtaattc 10620
atatggatat tcagaaaaaa ctggtgtgtt aaacactaca gatattgaaa gtgataaagt 10680
aacattaaat gttcctgcaa cagctgtata tcaagggata tcaaaaggaa atgctgattt 10740
aaatacacaa gtagatagat tatataatag tatgcttgca tgggaacaaa taatggattt 10800
agtttactca gaaagaggag tatttaaaac acaagattta gatggtaatg gaactgtaga 10860
tgataatgaa tttaacttaa ctaaaaatga tatggctcca aaatctcgta tgaacataaa 10920
ataccaaaga atgtttattg gagcatttat gtatgcttct ggattacatg taggagtaga 10980
atatggttca gttccagggc ttatgaatgg agtacctttc caaatagatg atagtggaaa 11040
agctacagga ggaaacttat ttggatgggg aataggtcat gagataggac atgttacaga 11100
tattggaaag atgacatata gcgaaactag taacaatgta cttgctttac ttgctcagac 11160
ctttgatgat aagacacatt caagacttga aggttcaaca atggataaga tatatgagca 11220
tgtaacttca aatagcttag gtataccaag caatgtattt gaaaggcttg gaatgttatg 11280
gcaattacat ctagcttatg atgatgattt tacgggttca atgttaaaga ataactcaga 11340

tgctgattta agtaatgata cattctatgc aaagatttca agaaaataca gagcactttc 11400
aagtagtgac ccaattaata gtttgccaaa agaccaaatg ctagtagcaa tggcatcaag 11460
tgttgttgaa aaagatttaa gagattactt taaggcatgg ggagtagaaa ttacaccaga 11520
attaaacagt ataatggata gtaagaacta cgaaaaagaa tcaagaaaga tacaatatgt 11580
aaatgacgaa gctagaagaa agattttaaa tgaaaatata gctagtatgg caaaagatac 11640
taaagtaagt gctagttttt ctgatggaat acaaagtggt tctctagtta aatctaataa 11700
aatatctata gacctagatg tagataagga taaagataag atactaggat acgaaataat 11760
aagaagtgat ggaaattatt tagatggtga aaatggaaca caggtaaaat acagaacagt 11820
tggatttgtg gatgcaaaag gaaatagtaa gactaaattt actgataata tatctcctct 11880
taataacaga gcatttacat ataaggtagt tgcatatgat tatcacctaa atccaacaga 11940
agaatttgaa gttggtactg taaaattatc agatgaaggg aaaataaata aatcagcatg 12000
gtcattcata acaaatactg tttctgacgg tgatgttaga actgaaaatg actcacatgg 12060
accattacaa aatcctgaaa ttgataatat aaaagataat gatgcaagta ctacatataa 12120
aggaaagata ataagtaaag cagagtggaa caaagaccct caaaaagacc ctgatataaa 12180
tgtagaagaa aatccatata taatagtaga catgaaagaa acattaccta tagtaggact 12240
aaaatataca aaaccagagg cagaagctag gaagttttca ttaaaaggat tatttaactt 12300
tagaaagaat actagcacaa catacaatcc tttaactaat tacaagatac aagtaagtaa 12360
tgataaaaag aactgggaag atgtaagtca aggaactttt gaatatggac aagatttagt 12420
tggtggagga aaagaccaag ataatgaagc tagagttaca tttaataaag atggtaaact 12480
ttggacatac caagctagat atgtaaaact aatctcaaat aataagtcta atatagaagt 12540
tgcagaatta gatataatag gacctccagg tgataatata gaaattggca caatgaataa 12600
tgcagaccaa actcgtacta atggaattgg aaaattagca gaagattatg tatatcaaag 12660
tgatgactct agcacaagtg aagatgaaac aggtaagatt ccagcaggtt caatagttat 12720
tacaggtgaa tacagtggaa atccagcatt taatttacca ctacttatag ataaaaacaa 12780
taagactata tctggagaag cacttttatt tgcagaagtt ccagaaaaag gtgaactagg 12840
tgaaatcagt aaaggtacat gggtttacta tatgccaaaa gaaaactttg attctttaag 12900
tgataaggtt aaggctgaac tatatagata caatgacctt caaggagata ctccagtagg 12960
tcaaagattt gtaagtgata caatgtatgt acctgtagga gctaagaatt atgatgagtt 13020
aaagacaata agtcttactg atagtaatag taaagctaga aaagctacaa gaaatgctat 13080
tgatatcagt aataagaagg ttataactgt aagcaataaa gtaaagggta atattgttag 13140
agataaatag taggtggaac tctaaaaacg taaaaatact aacactaaca ttgtatatct 13200
aataagatta ggtactagaa tagttttaaa atagtagata ttaaaagtaa agtatataaa 13260
tgtaactata agtttaaggt aatttataaa ttaaatataa aagctaggtg gataagtaaa 13320
tttatccact tagttaatga tagataagga cgggtgttaa tgaataaaag aaaatctttt 13380
ataagaacta tagcagtatc aactatggca gtagctgtaa ctggtagtgc tacatgtgca 13440
tatgcagctc cagttttaca aggaacaaaa acgtatgaga aagtaaatac tatagatata 13500
tctgtggact cagtagaaaa catagtatat tcatttcaag ccagtataaa agtacaaggt 13560
gaggttgaag tagtagataa tgagcagaaa gaaaagatta cttggtctga taatataaaa 13620
tctcagataa aaagtggaaa tgcaaatgct acttgtagag cagaatacaa caaatcaagt 13680
aatacaacta cattagatat atacgttaca tcaaatgaag atttactaga cggaaataga 13740
ttgaatatag gaagaatatc tgtaaaaaaa tctggcagta actcaaatgc tgactataaa 13800
gttttaggaa aaggaacaag tgataaacca gctttaaaaa ttgtaacata taacaacaaa 13860
acagtggatt atgaaaatat aagctctgat gaaggattaa tatttaccct tattaatgag 13920
agtgaagtaa aaccaattgg tggaactggt tctagtaaaa atgaccctga aaagtataag 13980
gtagagaaaa gtgaagcact tgaatattta ttaaacaata ttaggataaa ctattctatt 14040
gttagtaaag aaacacaaga aagtggttca aatgtaatat taaaacttgg attagcacaa 14100
aaaacaacta agggaagaaa agctacaata aacaaatatg ttgaagttac acttccaaaa 14160
agtttagaat acatagttga aaatgaatta tcaaaaccgg atgaattacc accagataat 14220
ggtagtggtg gaaataatgg tggaggaagt aattcaggcg gctcatctag tggtggaagc 14280
tctggaggag gaaattcaag tgattctaca tctaatgtca cagtcaaaaa attaaaaggt 14340
gctgatagat ttgagacagc tatcaaaata tctcaaagtg gttggacaaa atcagataca 14400
gttgtaatag taaatggaga agataaaagt atggtggatg ggcttacagc tacaccactt 14460
gctagtgtaa aaaactcacc gattctttta tcatcaaatg aaaagcttcc tcaaaaaact 14520
gtagaagaat tgaaaaggct aaatccatca aaagtagttg tgataggtgg aaacaactca 14580
atgccaaatt cagttgttga agctatcaaa gctgtaaatt caaaaatatc agtacaaaga 14640
ataggtggag atacaagata tcaaacttct attaatatag caaaagaaat agataggacg 14700
aataatgtga gcaagttgta tataggagca ggaaatggag aagctgattc attatcaata 14760
gcatcactgg ctggtaaaga aaaaactcca atagtactta cacaaaaaga tggtgtagat 14820
aatgaagctg aacaatttat taaatcaaat aaggtgtcaa atatatattt tattggtgga 14880
gtagaaaaaa tatctaacaa ggctattgaa caagttggaa aaatagcaaa caaagatata 14940
tcaaacaata gggtggcagg tcaaactaga caagagacga atgccaaagt aatagataaa 15000

ttctactctc aatctaagct agatggagtt gtagttgcaa atcaagacaa acttatagac 15060
gctttagcag ttggaccatt agcagctaaa aacaattcac cagttatact tgctacaaat 15120
acgttagata aatcacaaga atcgtcatta aaaggtaaaa attcatcaaa gttgtttgaa 15180
gtaggtggag gaatagcctc atctgtaata gacaagatta agagtttaat tgaaaaataa 15240
atattaagta aatagaaaat aagaatgtat ttagatataa agttatatat ttaaatgcat 15300
tcttatttta ttttcattat cttaacaatc taataatatt gaggttaaaa tgtataataa 15360
tttgagctat attaagttta attttaaata aagcagatat cttttatata tataaatata 15420
aattcaattt tgaatatgtt taaaaagaag ttaaaaataa ttttttaaat ttaataaaat 15480
aaaacactaa tatatagacg agatatattc atatcatata tactctataa tatatggaaa 15540
ataaagagaa aataaatgga aaataatcac aatataaata taaaatataa aaatcatcag 15600
ggggatatga aaatgaacaa aaaaatatta tcattaggtc tagcagtatc attaatctta 15660
gtaaacttca aaagtgtaaa tgcatcatca gtagtagaaa aaatatacgg taaagataga 15720
tatgaaacag cagcaaagat agctgataaa cagacttatg aaacagtaat tttagtaaat 15780
acagaaaaat cacttgcaga tggattaagt gtaagtggat tatcaggagc tacaaaagct 15840
ccaatactat ttacacaaca aaataagata ccagcagaca caaatagatg tctaaaaaat 15900
atcaaaaaag catatataat aggtacagaa gatactataa gtaaatcagt agaaaaagaa 15960
ctagattcta aaaatataga agtaaaaaga attggtggag aagatagact aaaaacaagc 16020
tatttaatag ctaaagaaat agcaactata aaaaaagtgg ataaagtact attaactaat 16080
gcatatagtg gagaagcaga tgcaatgagt gtatcatcag tagctactag agatggagct 16140
ccaatcatac ttacagatgg aaagagtgtg ccttttgatg taaaaaatat tcaatcatat 16200
tgtataggtt cagaggagat aatgagtaat cctctagtta aaaatacaaa ctcagtaaga 16260
atagaaggaa ctgaccgatt tgaaactaac aaaaacgtaa tagactactt ttttaatagt 16320
gcagatggat tctatgtatc agatggatac caattagtag atgcaatagc agctgcacca 16380
cttactaaaa actctccaat ggtactagta aatgatggaa gcgataaaac tgtattagaa 16440
ggagctaaga atataacctc tgtaggtgaa ataaatgaga aggtaataca acagtgtata 16500
aatgcttcaa agtcaaatgg acaaccccca acaattacag ttggaagtac agaggtatat 16560
aagggtgaaa agtttgacac tggcaagtta aatatagtag ctaaagataa cacagggaag 16620
gtattaccaa tagaagttga tgggtttata gatacaaata gagtaggtac atatatattg 16680
acattaaaag ctacagacga atggggaaaa agtacaggaa aaagagtaga aataaaagtg 16740
ttggatgaca aatctcatga ttacaacagc ccagaattta aaaaaatggt atctactgaa 16800
atgtataact taatcaattc ctatagaaaa gaaaaaggca aagagccttt ggtagtgtct 16860
agtaggctag aaggtatggc aaatgcatgg tctaagtata tgatggataa aaaagtattt 16920
gcacattata tagatggaaa aaatgctcca caagtatttt ctgagtttgg aatgagaagt 16980
gaggaaaata tagcatatat ttatatagat tctaagaacg ttcaaactac acaagatgca 17040
aaagacttgg caaaggctat atttgaggtg tggaagaaat ctccagagta caatgcaaac 17100
atgcttagtg atgagtttta tagtactgga tttgggcttt atatattatc tgatggtcaa 17160
gtgcatgcta ctcaagagtt tttaaatggc aatgaaggta gcttgtaaaa aattccataa 17220
atcttgtttt tttgttaatt tactgatatt ataaaatgaa aatcacatct aaagtacttt 17280
ttgtaatctg ttttgttatt gctttacagg ttatggtatt ttaggtgttt tttgttttag 17340
tagtttataa attttgaggg aattactact gaggatgtaa tttatgatta gctgaattat 17400
ataaaaaaat tagtaggagg agtttgatga aggtaaataa gagagtatta tcaataggat 17460
tggcaatatc attaataatg gcaggagcac caaatatcaa tgcactttca agtatagaaa 17520
agatacaagg gaaagataga tatgagacag ctgccaaaat tgcacagaaa caaacatatg 17580
aaaatgtagt attagtaaat acagataata cactagctga tggattaagt gcaagtggat 17640
tagcaggaac agttaaagca cctatattat tatcacaaag aaacagtata ccatcagaca 17700
cagaaaaaat gttaaaagat gtaaagaagg tttatataat aggaactgaa gattcaatag 17760
gaaaatcagt tgaaaatgaa ttaaagcaaa aagggataga agtcaaaaga ataggtggta 17820
atgatagaat tgaaacaagt tatcttatag caaaagaaat agcttcaata aaacctatag 17880
ataaggtttt tataactaat ggttatacag gtgaagcaga tgctatgagt gcgtcatcag 17940
tagcatcaag agatggagca cctataattt taactaatgg taaaaatgtt ccatttgaga 18000
aaaaagaagg tgtgcaatgt tatgctcttg gttcagagga gataataagt aatgatttag 18060
ttaaaaaaac taattcagtt agattagctg gagaagatag atttgagact aataaaaaag 18120
ttataaaaca tttttatagt tcagcaaaag aattttatct ttctaaaggt tatcaattag 18180
tagatgcagt tgctggttct tctatagcta aaaatgcacc tattgtattg gttgatggaa 18240
atagtgataa gtctgtactt agaagtgctg ataagattac agcactagga ggaatagatg 18300
aaaagacgct agaacaatgt ttaagtgcat catcactgga tgcaagtgca ccaacaataa 18360
ctgttggaaa tttaaatatt tatcaaggtg ataagtttga tataagtaaa ctgaatatcg 18420
tagcaaagga tagcaatggt aatgatttga caccagagct tataggaaat attaatactg 18480
ataaggttgg taaatacaaa gttactataa aagctactga tattggtgga aagactacct 18540
caataagtgt agaagtaaat gtacttgaat ataagactaa tgatatgaat agtagtgagt 18600
ttaaaagaat ggtttctagt gaaatgtata gtttggtaaa ttcatataga aaagaaaaag 18660

ggaaagagcc attacaagtg tctgaaaatt tacaaggaat ggcaaattat tggtctaagt 18720
atatggctga taaaggtgaa tttgcacatg taataaatgg taaaaatgct gctgaagtat 18780
tttctggtgg tattagaagt gaagaaaata tagcctttgt accacttact acaaagagta 18840
cttatacaac aaaagatgct agagaagtag caaatgtcat atttacagtt tggaagaagt 18900
ctgataagta taatgaaaat atgttgagtt ctaagtttgc atatactgga tttgggttat 18960
atatattacc agatggacaa gtgtatgcta ctcaggagtt tttaaataag taaaattaaa 19020
ttttttagtt tattacattt taaaatttag ggtataaaaa cttgtaaact tggagaaaat 19080
aataatttaa aaaaatagct tgcaaaaaga ataaaaatgg attattatag agatgtgaga 19140
aatattagga atatatggat gattattcta tgtacataat aaagagatgt aattttaata 19200
taatgttggg aggaatttaa gaaatgaata agaaaaatat agcaatagct atgtcaggtt 19260
taacagtttt agcttcggct gctcctgttt ttgctgcaac tactggaaca caaggttata 19320
ctgtagttaa aaacgactgg aaaaaagcag taaaacaatt acaagatgga ctaaaagata 19380
atagtatagg aaagataact gtatctttta atgatggggt tgtgggtgaa gtagctccta 19440
aaagtgctaa taagaaagcg gacagagatg ctgcagctga gaagttatat aatcttgtta 19500
acactcaatt agataaatta ggtgatggag attatgttga tttttctgta gattataatt 19560
tagaaaacaa aataataact aatcaagcag atgcagaagc aattgttaca aagttaaatt 19620
cacttaatga gaaaactctt attgatatag caactaaaga tacttttgga atggttagta 19680
aaacacaaga tagtgaaggt aaaaatgttg ctgcaacaaa ggcacttaaa gttaaagatg 19740
ttgctacatt tggtttgaag tctggtggaa gcgaagatac tggatatgtt gttgaaatga 19800
aagcaggagc tgtagaggat aagtatggta aagttggaga tagtacggca ggtattgcaa 19860
taaatcttcc tagtactgga cttgaatatg caggtaaagg aacaacaatt gattttaata 19920
aaactttaaa agttgatgta acaggtggtt caacacctag tgctgtagct gtaagtggtt 19980
ttgtaactaa agatgatact gatttagcaa aatcaggtac tataaatgta agagttataa 20040
atgcaaaaga agaatcaatt gatatagatg caagctcata tacatcagct gaaaatttag 20100
ctaaaagata tgtatttgat ccagatgaaa tttctgaagc atataaggca atagtagcat 20160
tacaaaatga tggtatagag tctaatttag ttcagttagt taatggaaaa tatcaagtga 20220
ttttttatcc agaaggtaaa agattagaaa ctaaatcagc aaatgataca atagctagtc 20280
aagatacacc agctaaagta gttataaaag ctaataaatt aaaagattta aaagattatg 20340
tagatgattt aaaaacatat aataatactt attcaaatgt tgtaacagta gcaggagaag 20400
atagaataga aactgctata gaattaagta gtaaatatta taattctgat gataaaaatg 20460
caataactga taaagcagtt aatgatatag tattagttgg atctacatct atagttgatg 20520
gtcttgttgc atcaccatta gcttcagaaa aaacagctcc attattatta acttcaaaag 20580
ataaattaga ttcatcagta aaatctgaaa taaagagagt tatgaactta aagagtgaca 20640
ctggtataaa tacttctaaa aaagtttatt tagctggtgg agttaattct atatctaaag 20700
atgtagaaaa tgaattgaaa aacatgggtc ttaaagttac tagattatca ggagaagaca 20760
gatacgaaac ttctttagca atagctgatg aaataggtct tgataatgat aaagcatttg 20820
tagttggtgg tactggatta gcagatgcta tgagtatagc tccagttgct tctcaactta 20880
aagatggaga tgctactcca atagtagttg tagatggaaa agcaaaagaa ataagtgatg 20940
atgctaagag tttcttagga acttctgatg ttgatataat aggtggaaaa aatagcgtat 21000
ctaaagagat tgaagagtca atagatagtg caactggaaa aactccagat agaataagtg 21060
gagatgatag acaagcaact aatgctgaag ttttaaaaga agatgattat ttcacagatg 21120
gtgaagttgt gaattacttt gttgcaaaag atggttctac taaagaagat caattagtag 21180
atgccttagc agcagcacca atagcaggta gatttaagga gtctccagct ccaatcatac 21240
tagctactga tactttatct tctgaccaaa atgtagctgt aagtaaagca gttcctaaag 21300
atggtggaac taacttagtt caagtaggta aaggtatagc ttcttcagtt ataaacaaaa 21360
tgaaagattt attagatatg taatataagt tttaataaaa ctttaaatag aaaaaggctt 21420
ctctcatgag aagtcttttt tatttaaaat aaatataaaa taaaatagag gctataaata 21480
gcctctattt tatgtgagaa atccctaaat aaaaagatgt ttaatttatt taagtaaaat 21540
cattatacta taaatcaaaa taattgtcta ctattatgga tttttaatgt ataatagtaa 21600
ttggacaata gaaaaggagg tacttatatg tcagttatag attcgatact tgacaaagca 21660
gatgagcaag aaataaagaa attgaatgtt atagtagata aaatagatgc attagaagat 21720
agtatgaaaa atctatcata tgaagaacta aaagacatga cagctatatt taaaaataga 21780
cttaaaaaag gtgaaacctt agatgatata cttccagaag catttgctgt agtaagagaa 21840
gtatcaaaaa gaaaattggg aatgcgtcaa tatagagttc aattaattgg tggaatagta 21900
atacatcaag gaaagatagc agagatgaaa actggtgaag gtaaaactct agttgaagta 21960
gcaccagtat atttaaatgc ccttacaggt aagggtgtac atgtaatcac agtaaatgat 22020
tacctggcag aacgtgataa ggagcttatg agcccagttt atgaatctct tggaatgaca 22080
gtaggagtaa ttatatctaa ccaagaccca aatataagaa aacaacaata taaatgtgat 22140
ataacttatg gtacaaatag tgaatttgga tttgattatt taagagataa tatggtgcca 22200
gatttatctc ataaagtaca aagggaacta aattttgcta tagtagatga agttgactca 22260
atattaatag atgaagccag aactccactt attattgcag gagatggtga tgaagattta 22320

aaactttatg agttagcaaa tagctttgta aaaactgtta aagaagaaga ctttgagctg 22380
gacagaaaag ataaaactat agcattaaca gcaagtggta taagtaaagc tgagtcattt 22440
tttggtataa caaaccttac tgatataaag aacatagaat tatatcatca tataaatcaa 22500
gctttaagag gtcataagct tatggaaaaa gatgttgact atgttatttc aaatggagaa 22560
gtaatgatag ttgacgaatt tacaggaaga gtaatggatg gtagaagata tacagatgga 22620
cttcaccaag ctatagaagc aaaagaaggt gttgagataa agaatgaatc taaaactatg 22680
gctactgtga cttatcaaaa tttcttcaga ctatatgaaa aactttctgg tatgactggt 22740
actgcaaaga cagaagaagg ggaatttgag tcaatctata aattaaatgt tgtccaaata 22800
ccaactaata gaccagttat tagagctgat ttacatgata aggtatttaa aacagaagaa 22860
gaaaagtata gtgctgttgt agaagaaata ataaggatac ataagactag acagccaata 22920
cttgtgggaa cagtttctgt tgaaaaatct gagaagctgt ctaaaatgct taaaaaacaa 22980
ggtattaagc atcaagtctt aaatgctaaa caacatgata aagaggcaga gataatttct 23040
aaagccggta aattagatgc tataacgatt gctacaaata tggcaggtag aggaacagat 23100
atttctctag gtgcaggaga taaagaagaa gaacaagaag ttaaagattt aggtggactt 23160
tatgttatag gaacagaaag acacgaatca agaagaattg ataatcagct tagagggcgt 23220
tctggtcgtc aaggagaccc aggtacatca agattctttg taagtcttga agatgatgta 23280
ataaagcttt atggtggaaa aactatagag aaacttatga agagaacaag ttcaaacgaa 23340
aatacggcta ttgaaagtaa agcacttaca agagctatag aaagagcgca aaaaggtgta 23400
gaaggtaaaa attttgaaat aagaaaaaat gttcttaaat atgatgatac tattaatgag 23460
caaagaaaag ttatatataa tgaaagaaat aaagtgttaa atgatgaaga tatacaagaa 23520
gatattcaaa aaatggttaa agacatcata caagaagcag gagaaactta tttaattgga 23580
agaaaaagag attattatgg atattttaaa catttatata gtacatttat gccagcagat 23640
acattactaa tacctggtgt agataaaaaa agtgtccaag agataattga tagcacatat 23700
gaaatttcaa aaagagttta tgacttgaaa aagatgatgc ttggtattga taaggttgca 23760
gagttagaaa aaacagtact tttaaaagtg gttgatcaat actggataga tcatatagat 23820
gctatggaac agttaaaaca gtatataggt cttaaatctt atgctcaaaa agacccattt 23880
aaggaatatg ctttagaagg atatgacatg tttgaagctt taaataaaaa tataagggaa 23940
gcaacagtgc aatacttata taaatttaac taaaaaagtt catgtaattt ttattaaatg 24000
aaaaataaag tattttgatt ctttatattg taattacaat taaataatta aggaatacaa 24060
atgaaaaaga agctgaaaat cagcttcttt ttctgtgtga gaaatacgga aaaaagagat 24120
gcttgaaatt tctttcttcc atttaatatt atactataaa gaagatgaaa agtctattag 24180
ttataaactg gattattaaa aaaaattaca attttgttac ttttttgtat acataaatat 24240
taaaaaaata aaatataaga aaaagtttat atcttttggt taattattac aataagtctc 24300
atttattgaa ataatatcaa atatatatta taattggtaa ataaggaaaa aataataaaa 24360
atttgaattt tttaggggga aaataccatg aataaaaaaa atctttctgt aattatggct 24420
gctgcaatga taagtacatc agtagctcca gtttttgctg cagaaactac acaggtaaaa 24480
aaagaaacaa taactaagaa agaagctaca gaactagttt cgaaagttag agatttaatg 24540
tctcaaaagt atacaggagg ttctcaagtt gggcaaccaa tatatgaaat aaaagttggt 24600
gagactttat caaaattaaa aataataact aatatagatg aattagagaa attagtaaat 24660
gctttgggag aaaataaaga acttattgta actataacag ataaagggca tataacaaat 24720
agtgcaaatg aagtagttgc agaagcaact gaaaaatatg aaaattcagc agacctttcc 24780
gctgaagcta actctataac agaaaaagct aaaactgaaa ctaatggaat ttataaagtt 24840
gcagatgtaa aagcttcata tgatagtgct aaagataagt tagttataac tttaagagat 24900
aaaacagaaa cagtaacttc taaaactata tatgtaggta ttggtgatga aaaagttgat 24960
ttaacagcaa atccagttga ttcaactgga acaaatttag acccttctgc agaaggattt 25020
agagtaaata aaatcgataa actaggtgta gcaggagcta aaaatattga tgatgtccaa 25080
ttggctgaaa taactataaa aaatagtgac ctaaatacag tttcaccaca agatttatat 25140
gatggatata gattaactgt taaaggtaat atggtagcaa atggtacatc aaagtcaatt 25200
agtgatattt cagcaaaaga ttcagaaaca ggaaaatata aatttactat taagtatact 25260
gatgcatctg gaaaagcaac agagcttact gtagagagta ctaatgaaaa agatttaaaa 25320
gatgccaaag ctgcattaga aggtaattca aaggttaaat tgatagctgg agatgataga 25380
tatgcaactg cagtggctat agcaaaacaa acaaaatata ctgacaacat agttatagtt 25440
aattcaaata aactagttga tggattagca gctacaccac ttgctcaatc taaaaaagca 25500
cctatattat tagcatctga taatgaaata ccaaaagtaa ctttagatta tataaaagat 25560
ataattaaga aaagcccatc agctaaaata tatatagtag gtggagaatc agcagtatca 25620
aatacagcta aaaagcaatt agaatcagta actaagaata ttgaaagact agctggagat 25680
gatagacata cgacttctgt agcagtagca aaagctatgg gttcttttaa agatgcattt 25740
gtagtaggtg cgaaagggga ggctgatgct atgagtatag ctgctaaagc tgctgaactt 25800
aaggctccta taatagtaaa tggttggaat gatctttcag cagacgctat caaattgatg 25860
gatggaaaag agattggtat agttggtggt tctaacaatg tatctagcca aattgaaaat 25920
caacttgccg atattgataa agatagaaaa gttcaaagag ttgaaggaga aacaagacac 25980

gatactaatg ctaaagttat agaaacatat tatggcaaat tagataaact atatatagca 26040
aaagatggat atggaaataa tggtatgcta gtagatgcat tagcagcagg acctctagca 26100
gcaggtaaag gtccaatact tctagctaaa actgatataa cagactcaca aaagaatgca 26160
cttagtaaaa aattaaatct tggtgcagaa gtaactcaaa taggtaatgg agttgaattg 26220
acagtaatac aaaagatagc taaaatacta ggttggtaat aacttatact gtggagaaga 26280
ttccttaatt aatacaggaa tctttttcta taattatgct atttattgta aattataaca 26340
ttttataaag taggaggatg tagtagtggt agatgcaact ttatttttta caccacatca 26400
agatgatgaa accttaagta tgggaagtgc tataattgaa catgttgaaa aaagtgatac 26460
acatgtcata ttatgcacag atggtagtaa atctataata agaaaagttt tggatgatgg 26520
agggtcatgt tcttatcata ttaaagatgt tcataaatac tctttatcag aaagtgaatt 26580
ttcaaaagat agagatgaag aatttaagga tagctgtgaa gcaatgggtg taaaagaaag 26640
taatatacat atagaagata atagagcaca tgatggtgag ttaagtaaag aaaaggcaag 26700
ggaaataata cttaagtact tagaagaata tccagatgca aaagtaaaaa ctgttactcc 26760
atttaaagct agtggaatac atgaagacca tagagcatta ggagaagctg cattagagtt 26820
atatagagaa ggcaaaataa aagatttaag attttatgta gaaccatacg actataagga 26880
ttttaaaaaa gttaatccaa atgtagaagt atggaaggta ttacctagtc aagaggagaa 26940
gttattaagt gcaatgaatg catataaaaa atggaatcct gaaagtggac attatgctat 27000
aggatatcat tcagtaaaat ctcatttga tgagttagct acaaataaaa tacagtatgt 27060
acatgctcca taaaatatta cgaaagaatt aggaggtaag aattaatgaa aatatcaaaa 27120
aagatagtgt ctttgttaac tatgacattt ttaactgtta cattatatgg aaatacatct 27180
aatgcatcta caaaagatac attaacgggt tctggaagat gggaaacagc aataaaaata 27240
agtcaagctg ggtggacaaa gtctgaaagt gctgtcttag taaatgacaa ttccatagca 27300
gatgctttat cagctactcc atttgctaaa gcaaaagatg cgccaatatt attaactcaa 27360
agtaataaat tggatagtag aacaaaagca gaattaaaaa gacttggtgt gaaaaatgta 27420
tatttaatag gtggttcaat tgcattaagt tcagagattg aaaagcaatt aaatgcagaa 27480
aatataaatt ttgaaagaat atccggaaat agtagatatg atacttcttt aaaattagct 27540
gaaaagctag atagggaaaa gtctatatct aaaatagtag tagtaaatgg agaaaagggt 27600
cttgctgatg cagtgagtgt tggagctata gctgctcaag aaaacatgcc aataatactt 27660
tctgattcag agaatggaac tgaagtagct gataatttta tagatagtaa agatatagca 27720
aaatcgtatg taataggtgg tacatattct atttctaatt ctgtagaaag aagtttacca 27780
aatgcaacaa gaatagcagg tagtagtaga agtgagacaa atgcaaaaat tatagaagaa 27840
ttttataaag atactgatat aaaaaatatt tatgtaacaa aagatggtac aaagaataag 27900
aatgatttaa tagattcttt agcagtgggt gtattagcag ctaaaaatag ttctccaata 27960
gtattagcag gaaataagct tgatactact caaaaagatg ttttaaatac taagattata 28020
gataaagtta ctcaaattgg tggcttaggt aatgaaaatg ttgtagaaga tatactagat 28080
atacaagaag agactaagta tactgttgaa actattgatg aacttaatgc tgctataaaa 28140
agagcagatg caaatgatat tataaagttt aaaccagaga aagaaaaaac tataaataat 28200
tcttttagta ttgaaactaa aaaaactgtt actattgaat tagatggaag atatagacaa 28260
acaataactt tggatatacc taatggaaaa tttaataact atgcagaaat agaaggtgga 28320
gtaaagctaa aaaacataaa aaatgaatca ttagttaata aaggaagtat acaggattta 28380
gatatatatg atgaaaatgg ttgtaaaata gagaatgaaa gttctggaga aatttggttt 28440
gttactatag ttgaagaagc taatgatgta tatatagtta atagtggtga tataacaaaa 28500
atatcaaata attcttctag tactataata agaaattctg gaaatattga tacagttaca 28560
ggaaaaaaag aacctgcaat aagtggaaat aaacctaaag tcaatgatac agaaaaagaa 28620
actaaagctg ctagaggctt aaatcctagg gtagaagcat gttctgttcc taaaaaagat 28680
tatgttatga taacaattcc taattcacct aaagattcaa gatataaaat ttactataga 28740
gtagtttata ataaacctta tgcaatggat gttggtgata aaattaatat tggagaatgg 28800
actgttgcac caacagatga agagccattt cttgaaaaag ctaaaaatgg ttgttatgta 28860
gaagctgttg aggtaaatac ttcaacgaaa gaggtttcta gatggggaag aactaacgct 28920
acagatgatg gattttaaat attaaaatct tttaaaaaca aaaggtagac aatttttttga 28980
ttgtctactt tttatttta aaaattatca ttttactaaa caaataattt aaaactataa 29040
caattatatt agtaaatact tttacaataa catcatttac aaaaagcaaa tcaatcatta 29100
catacataat taacatttca acgacaccag aaagtagcct aaaactaaca aaagacaaaa 29160
attccttaaa taagaacttt atttcaattc ttttactttc aaactcaaaa aacttattag 29220
taacataagc aaataagaca gcaagtatcc atgccaaagc atttgcaacc ataaaattaa 29280
ataaaataac tctagtaaaa tataaatacg aaactatatt tacaagagta gtaaatgctc 29340
caaagaataa atataatatt gtctctttat gttttttaa aatcaaatat aatcctccta 29400
agctaattta atataaatat acttaaagat actaacaaac attagtatag cacatctatt 29460
aataaatcaa tatatcttaa aatattataa aaataaaact ataatattct ataattaata 29520
tgtactcata atccatattt tatcaaaata aaatatgacc ttaatttaga aatctgatat 29580
aataaaaatt aaatacatat aaggggtaa acatgagaaa gtataaatca aaaaaattgt 29640

caaagctact agcattgtcg acagtttgct ttttaatagt gtcaacaata cctgtctcag 29700
cagaaaacca taaaactcta gatggagtag aaactgcaga gtattctgaa agctatcttc 29760
aatacctaga agatgtcaaa aatggagaca cagcaaaata taatggagta ataccttttc 29820
cacatgaaat ggaaggtaca acacttcgaa ataagggtag aagtagtctt ccatcagcat 29880
ataaatcaag tgtagcatac aacccaatgg atcttggtct tactacacca gcaaaaaatc 29940
aaggaagtct taatacatgc tggtcttttt caggtatgtc aactttagaa gcatatctta 30000
aattaaaagg atatggaacg tatgacttat cagaagaaca tttaagatgg tgggcaactg 30060
gtggaaaata tggatggaac ttagatgata tgtcagggag ttcaaatgta acagccatag 30120
gatatctgac agcatgggca ggtcctaagt tagaaaagga cataccatat aatcttaaat 30180
ctgaggcaca aggtgcaact aaaccttcaa atatggatac tgcacctact caatttaatg 30240
taacagatgt tgttcgtctt aataaagata aggaaactgt aaaaaatgct ataatgcaat 30300
atggttcagt gacatcaggt tatgcacatt attcaacata ttttaataaa gatgaaacag 30360
catataactg tacaaataag agggctccat taaatcacgc tgtagcgata gtaggatggg 30420
atgataatta ttcaaaagat aactttgcat ctgatgtaaa accagaatca aatggagcat 30480
ggttagtaaa aagtagttgg ggagagttta attctatgaa aggattcttc tggatttctt 30540
atgaagataa aactctacta acagatacag ataactatgc aatgaaatca gtatcaaaac 30600
cagatagtga taaaaaaatg taccaacttg aatatgctgg tcttagcaag ataatgtcaa 30660
ataaagtaac agcagcaaat gtatttgatt ttagcagaga ttctgaaaaa cttgactctg 30720
ttatgtttga aacagattct gtaggagcaa aatatgaagt atattatgca ccagtagtaa 30780
atggagttcc tcaaaacaat tcaatgacaa aacttgcaag tggaacagta tcatattctg 30840
gatacataaa tgtacctact aattcttaca gcttaccaaa aggtaaagga gcaatagtag 30900
tagttataga caacacagca aatcctaata gagaaaaatc aactttagca tatgaaacta 30960
atatagatgc atattattta tatgaggcta aagcaaactt aggtgaaagc tacatacttc 31020
aaaacaataa gtttgaagac ataaatacat atagtgaatt ttctccttgt aactttgtta 31080
taaaagctat aacaaaaaca tcttctggac aagctacttc aggagaatct ttaactggag 31140
cagatagata tgaaacagca gttaaagtta gtcaaaaagg atggacttct tcacaaaatg 31200
cagtattagt aaatggagac gcaatagtgg atgctttaac agctacacca tttacagcag 31260
caatcgactc tccaattctt ttgacaggaa aagataactt agattcaaaa actaaggcag 31320
agttacaaag attaggaact aaaaaagttt atctaatagg tggagaaaat tctttaagca 31380
agaatgtaca aactcaactt agtaatatgg gtatatcagt agaaagaatt tcaggtagtg 31440
atagatacaa gaccagtata tctctagctc aaaagttaaa tagtataaaa tctgtttcac 31500
aagttgcagt ggcaaatggt gtaaatggac ttgcagatgc aataagtgtt ggtgcagcag 31560
ctgctgataa caatatgcca ataatactta ctaatgaaaa gagtgagttg caaggtgctg 31620
atgagttttt aaattcatct aaaataacta agtcttatat aattggtggt acagctactt 31680
tatcatcaaa tttagaaagt aagctttcaa atccaacaag acttgcagga agtaatagaa 31740
atgaaactaa tgctaagata atagataaat tctaccctag ctcagatttg aaatatgctt 31800
ttgttgttaa agatggttca aagagtcaag gagacttgat agatggatta gcagtaggag 31860
cattaggtgc taaaactgat tcaccagtag ttctagttgg aaataagtta gatgaaagtc 31920
aaaaaaatgt acttaagtct aagaaaatag aaactcctat tagagttggt ggaaatggaa 31980
atgaaagtgc ttttaatgaa ctaaatactc ttttaggaaa atagtgaatg gaaagatagt 32040
aaggatatac attttattat taaaaaatgt tcaatagaag tttataaata taaaataaca 32100
aaaaggggtg tctcaaaatg aacttttag ttcatgaggc actctttttt gttgttttat 32160
atttttatca tttaaataat gaaaaatagg cttgtctata tttttagata gttttcttat 32220
tattgatgaa aaatgtaata atttttaaact gtaaaatagt gtaaagttat tttataatat 32280
aaatatagta agattttaaa caaaataagg gggaataata atgaaagcac caaaaactat 32340
tttaacaata ctgacaatag cgcttacttt aagtagcatt tctataatac catcatatgc 32400
acttacagag gaaaaattaa taggtaatgg tagatatgaa actgctgtaa agataagtca 32460
aaaggcatac agttctagtc ctaatgtagt gctagttaat gacaattcat tagctgatgc 32520
actttcagct acaccatttg caaaggctaa aggagcacca atacttttga cagaaagtga 32580
taagctagat gataggactg aaaaagagat aaaacgtcta ggagctaaag atatttattt 32640
gataggaggt actgctgtac ttaataaaga tatagagaac aagttaaaag gtaatggact 32700
taatgtagaa agaattaatg gtaagaatag atatgaaact tcactgattt tggctaataa 32760
gcttaaggat ataaaagata ttaaggaagt tgctgtagta aatggtgaaa aaggattaag 32820
tgatgctgtt agtgttggag caccagcagc tcaaaataag atgcctataa ttttgtctaa 32880
ccctaaggac ggagtagaag catttgataa atttataagg gatgagaaag ttataaaagc 32940
atatgtaatt ggaggtacaa attctgtttc aagagctgta gaaaagagtc ttcctaatgc 33000
agaaagagtg agtgggaaag atagaaatga gactaatgct aaagttatag aaaagtttta 33060
cactgataca aatttaagta atttatatgt aactaaggat ggtagtaaaa atgagaatca 33120
actaatagac tcgttagctg ttggagtatt ggctgcaaaa aatgagtctc ctattgtttt 33180
ggttggcaat aaattaaaca caaaacaaag agatatactt agtactaaaa aattaaatac 33240
tataacacag gtaggtggaa atggaaatga agaagctttt gatgaaatta aaagtctaca 33300

agaaaaaact gtgtttgaag cgaaaactgt tgaagaatta actgatatga taaatatagc 33360
aagtcctaat gatattataa actttaaacc aaaggagaat actgtaaatg aagcttttag 33420
aatggtaaca aataaaccaa taactgtgaa tataaagggt gattgttcaa agactcttac 33480
agtagatatg cctaatggtg aagtgaataa ctatgctact ttggtaaatg ttatagttag 33540
aaatattgga gaaggtggat ttaataatca tgatacaatc actatattgt cagttcgtga 33600
caaaaatggt agagttatag aaaatactag aaattcagat atagatacgt tgatgatatt 33660
agcttctgct aatgatacaa aattgattaa tgatggatat ataggaaaac ttatagataa 33720
ttcatctaat agtgatataa ctaataatgg tactatagat aaaaaagtta atcaggttga 33780
agatttagag gctaaagtag attcaattga aaaagctata gattctataa gtcaaaaagt 33840
caataaaata caagatatac tagataagct tggatttcta aagaaatttc ttagctaaga 33900
atatatatat ttaagtgaag tgtgtgttgt aattgttcta aagttatgta gttaaacagg 33960
tttatataca aatagattta tagataaata tgtagtttag tatatatttg aataagcttt 34020
ggggaggtaa catggggatt ggaattgaat gtaacaaaaa aataaataaa agtattgaca 34080
aaactagaat aaagaaaaat aatctaaata gcaatagcaa tagcaatagc aatagcaata 34140
gcaatagcaa tagcaatagc aatagcaata gcaatagcaa tagcaatagc aatagcaata 34200
gcaataaaat ttcaaataat gctgtagatt ctaaacttat gaatagtagg attaaatcaa 34260
tagatataat aagagggcta tcaatagccc tcatgatagt ctgcaataat cctgggactt 34320
ggatgagaat gtatcctcaa ttaagacatg ctgtatggca tggtgtaaca cttgcagatt 34380
ttgcatttcc ttttttgtt atatctttag gtgttacaat tccaatatca ataaatagta 34440
aactaaaaaa taataaatcc acactgagta taatacttag tatattcaag agaagtatct 34500
tattgatatt atttggattt ttcttaaatt acctaggaaa tcctgattta gatactgtaa 34560
gaatacttgg agtacttcaa cgtatgggat tagtctattt tgtgactagt ttagtatatt 34620
tactactaaa aaaattaaat gttggaagta ctgctacaat aattactttt ctatgtatat 34680
ctacatttat aatagttggg tactacatat tagcaaaacc ttatggattt gagctagaag 34740
gttctcttgc acagttagtg gatttacatt tcttcaaggg acatttgtac aagccagaat 34800
ttgaaccaga tggtttttta actagtatag tagcaatttc ttcaggtatg ttgggatgca 34860
ctatgggatg tgtactttta aaagaagaca taggagaata taaaaaattc tttaaaatac 34920
tggttatgag cataatcctg ctaataggtg cttttatctt taatcagtat tttccattta 34980
ataaaagatt atggtctagc tcatttgtac ttttaatggc tggcagttat ggtatattac 35040
tgtcaatatt ttatttata tgcgacatca aaaataaaag taaaatattt accccaataa 35100
tagcacttgg aagtagccct atttttacat atatgtgcct tgaaatatta agtcatgttt 35160
tttggaatgt tccaaaactt acgaacaaag tcgattatcc aactaccttg gtagagtgga 35220
ctacatatga acttataaca ccttgggcag gtacaacttg ggattctctt atattttctt 35280
tattgtatgt cctattctgg attatagtaa tgtctataat gtacaagaaa aaaatattta 35340
ttaaaattta gtgcaatatt tttagatgat aaagtctttg atttgttcaa atatacggaa 35400
actatcaaat ataatcaaaa tttctcctcg ataaaattta aaaaccatgc cataaaagat 35460
tacaaaaagt ttacaatttg taatttttg tgctataatg aaaaaagatt tacataatat 35520
tgaatttatg taatataaaa caagcttaag atgtattaaa aaagatttaa agagttaaag 35580
aaccaaaaaa tgaggggga tatgatgaaa aaaacaacaa aattattggc aacaggaatg 35640
ttgtctgtag caatggtggc accaaacgta gctttagcgg ctgaaaatac tactgcaaat 35700
acagaatcta attcagatat taatataaat ttacaaagaa aaagtgtagt gttaggaagt 35760
aaatctaatg caagtgtaaa atttaaagaa aaattaaatg ctgattcaat aacattaaat 35820
tttatgtgtt atgatatgcc acttgaagct actttaaatt acaatgaaaa gacagattca 35880
tatgaaggtg taataaacta taacaaagac ccagaatact taaatgtatg ggaacttcaa 35940
agcataaaaa taaatggtaa agatgaacag aaagttttaa acaaagaaga cttagagtct 36000
atgggattaa atcttaagga ttatgatgtt acacaagagt ttataataag tgatgcaaat 36060
tcaacaaaag cagtaaatga atatatgaga aaaacttctg caccagtaaa aaaacttgct 36120
ggagctacta gatttgaaac tgctgttgag ataagtaaac agggttggaa agatgggtct 36180
agtaaggttg taatagttaa tggagaatta gcagcagatg gaattacagc tactccactt 36240
gcatctacat atgacgcacc aatattactt gcaaataaag atgatatacc agaaagtaca 36300
aaagcagagt taaaacgtct aaatccaagt gatgtaataa ttataggtga tgatggttca 36360
gtttcacaaa aggcagtgtc tcaaattaaa tctgctgtaa atgtaaatgt gactcgtatt 36420
ggtggagttg atagacatga aacatcttta cttatagcaa aagaaataga taagtatcat 36480
gatgttaata aaatatacat agcaaatgga tatgctggtg aatatgatgc tcttaatatt 36540
tcatcaaaag ctggtgaaga ccaacaacct ataatattag ctaataaaga ttctgtacca 36600
caaggtacat ataattggtt atcttcacaa ggtttagaag aagcttatta cataggtggt 36660
agtcaaagtt tatcatcaaa aattatagac caaatatcta aaatagcaaa aaatggtact 36720
tcaaaaaata gagtttctgg agcagataga catgaaacaa atgcaaatgt aattaagact 36780
ttctatccag ataaagagtt gagtgcaatg ttagttgcaa aatctgacat aatagttgat 36840
tcaataacag caggaccttt agcagctaaa ttaaaagctc caatacttat aactccaaaa 36900
acttatgttt ctgcatacca tagtacaaac ttatcagaaa aaacagcaga aacagtatat 36960

caaattggtg atggaatgaa agatagtgtt ataaattcta tagcatctag tttatcaaaa 37020
cacaatgcac caacagaacc agataattca ggttcagcag caggaaaaac tgtagtgata 37080
gacccaggtc atgggggttc tgattcaggg gcaacaagtg gattaaatgg tggagcacaa 37140
gaaaagaaat atactcttaa tacagcatta gcaactactg aatatttacg ttcaaaaggt 37200
ataaatgttg ttatgactag agatacagat aaaacaatgg ctttaggaga aagaacagca 37260
ttatcaaata ctataaagcc agatttattt acaagtatac attacaatgc gtctaatgga 37320
tctggtaatg gtgtggaaat ttattacaaa gtaaaagata aaaatggtgg aacaactaag 37380
actgcagcat caaatatctt aaaaagaata ctagaaaaat ttaacatgaa aaatagagga 37440
atcaagacaa gaacacttga taatggaaaa gattacctat atgtgttaag aaataacaat 37500
tatccagcaa tacttgttga atgtgcattt attgacaata agagcgatat ggataagtta 37560
aatactgcag aaaaagtaaa aacaatggga actcaaattg gaataggaat agaagataca 37620
gtaaaataac atagataata ttaaagcaaa acatatcatc tcctgttata atgtaaaata 37680
taacaagaga tgatatattt ttgtaaagaa tcaaattgtt tacaatgatg tttcttgtat 37740
tgataatctg ctcattattg agtataatat agaaataagt caaaaataag atagaaataa 37800
gttgaaatta agaattaaaa aaatttttat tccaatatgg aggtaactat gtctggatat 37860
actaatgatg agtgtgaaat acctaaaatt aaaagctatc caggggcaga taaagaaatt 37920
gcaagtgaga ttgactatag catagtaaaa ggaacagttt tattcgactt gtaccaaaga 37980
attatggatt tagttttatc tataatagga cttgtgatag gtttaccttt aatagcaata 38040
tttggaattc ttattaaaat agaggacaaa ggaccaataa cttataagca agaaagacta 38100
ggaaaatgcg ggagaagatt ctatatatat aaattgagat ctatgagaac agacgcagaa 38160
aaatttggcg cacaatgggc tgaaaaagat gatcctagaa taactaaggt tggaaagttt 38220
attaggaaaa ctagaattga tgagattcca cagctgttta atatattaaa gggggacatg 38280
ggcttaatag gtcctagacc agagagaccg aactttactg ttcaatttaa tgaggaaata 38340
cctggattta tcaacagact cgctataaaa cctgggctta caggttgggc acaagtaaat 38400
ggtggatacg aaatcactcc tgaagaaaaa cttaaagaag atatttacta tataaaaaat 38460
aggagtatat tacttgactt taagatactc tttaaaactg ttaaagtagt attgacagga 38520
gacggagcta gataatttta aaatattaga ttactaatca atgacaaagc tctctaaaaa 38580
gagctttttt attgtcaaaa aatacaaaaa agaaaaggaa tgagaactta tgattaaaaa 38640
aatctctact atattatcgt tagttttgtt aatctctatt tctagtacga taggagtatt 38700
tgcagatgca aatcccaaaa gagagttgat agaagggagt ataccagaga tttcgacaga 38760
attaaataaa agagcattta aggattctaa agaagtaatt cttgtaaatg aagagtcaat 38820
tgtagatagt ataagtgcaa caccttagc atattcaaaa aatgctccaa tagttgtaac 38880
taagagtaaa aatcttggta gagtaacaag aaattatcta aaagaactag gtccagagaa 38940
ggttacgata gtaggtggtt taaaggcagt atctaaagat gctgaacgta atatagaaaa 39000
aatgggaatg aaagtagaaa gaataagagg gaaagacaga tatgatacat ctttaaagat 39060
agctagagaa atgtatagaa cagtaggatt tgatgaagca ttttttactta gctcaacaac 39120
aggattagag aacgccatat ctgtatattc ttatgccgct aaaagtggaa tgccaataat 39180
atgggctaaa gatgaagggt ttgaggaaca aattgatttc ttaaaaggca agaatcttaa 39240
aaaaatatat gcattaggtg attcaaaaga atttattgca gagattgatt ctaacttaaa 39300
aaatatagaa ggaataaagc aaataaataa gtcaagtaca aatgttgatt taataaagaa 39360
gttttatgat gaaaaagata taaagaaaat ctatactgct agattagact ttggtagtag 39420
atctgatgta aacgaatata tttctttggg agtagtatca gcaaaagaaa atatgccaat 39480
actgatttgt agtgataact taagtcgtgc acaagataaa ttcttgaaag atagcaatat 39540
aaatgatgta gttgaagttg ggtatacagt aggtgattat agtctattta agagcatatt 39600
taatctaaca ttcttatctt gtatagtatt aatactatta ttactattga ttacatttag 39660
ggcacttaga tatgagtcga aatagttttt gatttataag gaagcttttc ataggttaaa 39720
tttacttaat tagaataatg tttataagca gttttaaaag ttctactttt agagtttaaa 39780
ttagataagg agaaagaaga tgtcaaaaac agctaaagca gcattgtgga ttatggcagc 39840
cactatgttt tcaaaagtat taggtttcct tagagaactt gtacttgcaa atttctatgg 39900
gacaggtatg tatgcagacg tatttgtatt aacattaaat atacctggtt taatcatagc 39960
agtcataggc tctgcagttg ctacaactta tatacctatg tactttgaga caaagaaaag 40020
actcggagat gaaggtgcct taaaatttac taataatgtt ttaaatatat gctatataat 40080
ggctatagta atagctatta ttggtctttt atttacagag caatttgtta cagtatttgc 40140
agcaggattt agaaacgacc ctgctaagtt ccaagcagca atattattta ctaaaataat 40200
gatttcagga gttttattcc tgagtggaag taagatattt agttcgtttt tacaggtgaa 40260
tgacagtttt gtaattcctg ggcttatagg aatcccatat aatattataa taatagcagc 40320
aatagcacta agtgcaggta aaaatgtatg gatactgcca gcaggagcat tattggctat 40380
ggcgagtcag ctattgtttc aactgccatt tgcgtttaag aagtcctaca agtataagcc 40440
atatataaat ttaaaggatg agtctataaa agagttagtc aacctagtat tgcctatgtt 40500
ggtaggagta gctgttggac agttaaatat atttgttgac cgattattgg caaccacttt 40560
gggtgatgga aagttgtctg ctcttaacta tgcaaatagg ctaaatgagt ttgttatggc 40620

attatttgtt acatctataa taacagttat atatcctaaa ctagcaaaaa tgtctggtaa 40680
agataataaa gaaggattta taagcactat tgtaaaatct tcaaattgta ttatattggt 40740
ggtgcttcct atatctatag gagctattat attagcagag cctcttgtta gaatattatt 40800
ccaaaggggt aaatttgatg cgttatcaac agaccttaca tctatagcgc ttagattata 40860
ttcactaggt ttattggctt gtggagtaag agatgtctta tatagggcat tttactctct 40920
ttcagataca aaaacgccaa tgataaatgg aagtatagct ttaataataa atatagttct 40980
taatttaata ttgataagac cattgggtca tgcaggtatt gctatatcaa ctagtacatc 41040
aaacataatt acagttatat tattatttat atctcttaaa aagaaaaatg gatattttgg 41100
aggggataaa attataaaga ctggtctaaa aagtttggta gcatctggtg tgatggcagt 41160
agctacttta ttgatatata ataatctata tgcatttatg ggttcaggaa caattaaaga 41220
aattatatca gttggagctg gagtattagg tggagcttct gtatatactg tattgatagt 41280
tttatttaag gtagaagaga tggatttagc ttttgaattt ttaagaaaag gaaaacaaaa 41340
actacttaga agatagataa ttaccttgaa attttaaatt agagtcagaa aggatgaaaa 41400
tatggattat aagaataatt atgagatgtg gttaaattct ccttattttg atgaacaaac 41460
gaaaaatgag ctattaagta taaaagatga tgaaaaagaa atacaagata gattttataa 41520
aaatttagag tttggaacag gtggtttaag aggtataata ggagcgggaa ctaatagaat 41580
caatatttat actgttagac gtgcgacact gggcgtttta aattatatca tgaaaactca 41640
aggagaagaa ggtaagcaaa agggaatagt tatagcacat gatagtagat acatgtctag 41700
ggagttttgt atagaagtag caaagacttt aagtgcttat ggtgtaaagg cttatatttt 41760
tgaagaatta aagcctacac cagaactttc ttttgctgtt agatatttaa agtgtgctat 41820
gggaatagta ataacagcaa gtcataatcc taaagaatac aatggatata aagtatatga 41880
ctcagatggt ggtcaaatct gtatagatat ggcaaatgat ataatagcag aagttaataa 41940
gattgatgat tacagtacta ttaaaagtat agattttgaa gaagctttat ctaaaaattt 42000
aataactata ttagataatg aagtagatga tgagtttata aaagctgtta agaaacaagt 42060
tttaagacaa aatataatag atgaatacgg gaaaaaatta aaaataatat atacacctat 42120
tcatggtaca ggaaataagc cagtaagaaa agtcttaaat gaatgtggtt ttgaaaatgt 42180
aatggtagtt aaggagcaag aattgccaga ttctaatttt tcaacagtaa agtatcctaa 42240
tcctgaagaa aagtcagtgt ttaatatagc tattgaaatg gcaaagaata atggcactga 42300
tttgataata ggaacagacc ctgattgtga tagagtaggt atcgttgtaa aagattcaag 42360
tggtgaatat gtagtcttaa atggaaatca agtcggttca cttttagtta gatatatatt 42420
agagagttta gttgaagaaa ataaacttcc taaaaacaat cctacaataa taaaaactat 42480
agttacatca gaactaggtg caaaaatagc aaaggcttat aatgtagatt gtttaaatac 42540
tttaacagga tttaaattta taggagaaaa aataaaagcg tttgaagaaa gcaatgatag 42600
aagttttata atgggatatg aggaaagcta tggttactta attggaactc atgctagaga 42660
taaggatggt gttgtatctt cacttatgat atgtgaaatg gcagcatact attcttcaaa 42720
aggtatgaac ttgtatgaag ctttgataga tacatataat aaatttggat actacaaaga 42780
ggacttaaag tcagtaactt taaaaggcat agatggaata aaaaaaataa aagagatgat 42840
gctttacttt agaagtgtta aaatagataa tgtggctgat gtaaaagtag ataaaatttt 42900
agactataaa gatggtgttg atgatttacc taagtcagat gtacttaaat ttttacttga 42960
agatggttct tggatagcta taagaccttc aggtacagaa cctaaaatta aattttactt 43020
tggagcaaat tcagacaatc aagaagatgt agaatttaag ttgaacaatt taatatctta 43080
tatattaaat gtagtggatt ctatataaaa agctaagggg agataaacaa tgaaagtata 43140
caatgtcata atggctggag gaggaggaac tcgtttctgg ccacttagta gacaagaagt 43200
acctaaacag cttataaatt taagtggaga agatgcttta ataaacgaaa ctataaacag 43260
aatagattct ctagcaaaaa aagatgattt atttatagta accaatgaga agcaattgga 43320
agctttaaag gatatagtaa aagataaatg tttggatagt aatatattgc cagagccatg 43380
tgccagaaat acagcagcag ctataggttt tgcagcattt aacataatga aaaaatatgg 43440
tgatggtgtt atgtgtgtat atccagctga ccattatata aaagatgaaa aagaatttaa 43500
atctattttg gaaaaagcta tctatatagc agaaaataat gataaactag taactattgg 43560
gataactcct acatttccat caacaggata tggatacata aattttaata gagaaaatac 43620
tatagaagat gttgcatatg aggtagtgga gtttgtagaa aaaccaaact atgaaattgc 43680
taaagaatat gtaaactcta aaaagtatgt atggaatagt ggaatgttcg tatggaaggt 43740
atctaaaata ctagaagatt ttaagagata cttgccaaaa gtatatgaaa aattagaaga 43800
tataagtaaa tacttaggaa caaaagaaga aatggaaaaa ataaaagaaa tatacccaac 43860
tatacaatca atctctatag attatggtat tatggaaaga agtaatgatg ttatagtagt 43920
tcctggtgat tttggatgga atgatgttgg ttcttgggat tctctaggtg caatatatcc 43980
aactgatgat gaaggtaata tcaaaagagg tgaaaatatt actatagata ctaaaaattc 44040
aattatatat tcagatgaca aattaatatc gacaataggt attagtgatt tgatagtagt 44100
atctacaaat gatgcagtta tggtttgtag aaaagataaa gcctcaagat gtcataaaaa 44160
tagtagagca attaaaagaa gaagatagac aagaatatat gtaggaggtg ttcatttgat 44220
tcctaaaaaa atacattatg tatggtttgg tggtccaaaa ggaaatatag aaaatatatg 44280

tataaattct tggaaagaaa agcttccaga atatgagata gtagagtgga atgaaaaaaa 44340
ctttgatata gagaaagaaa taaaaggaaa taagtttttg gaagaatgtt ataaaagaaa 44400
actttgggca ttcatatctg attatacaag aataaaagta ctctatgaac aaggtggagt 44460
ttatatggat acagatatgc aaatattaaa ggatataact cctcttttag aaaataatag 44520
gctaatctgt gggtatgaag atgatagaga gtatataaat ggtgccataa taggtgtaga 44580
aaaaggacat cctttttaa aagacttatt ggagtattat gaaaaagaag tacttacttc 44640
ttcattgttt acaataccaa agataatgac tcatcttatg gaaaagaatt ataagaagat 44700
tgacccaaat aattatgaag aaggaatacg tgtttacgac aaagaatatt tttatccttt 44760
tggatttaaa gaagacttta caccagaatg tataacagaa aacacttttg ggatacattg 44820
gtggggaaaa agttgggcta aaaagagaaa ttatttctta gaaagcaaac acttaacagg 44880
tgtaaataag atatggaagt gttgtaaaat atttgcaagt aatactttac gaagctaaaa 44940
gggagaaagt aaaatgaaga ctttaaatga tgatagaaat attatgatta gaaaatatct 45000
ttatatattt gttttgttta tcatattact tagaaagtat atgatatcat ttattgaccc 45060
aaatatggat atagggatga ttaaatcagc gttattttat tcttcaatag gaatactaat 45120
gctacttttt ttatttgata agagaaaaag tataacagaa atggtactag ttggagtttg 45180
tgtcttactt tatttattaa atagggaagg agctatatta ttaatcgttt tattagctgt 45240
atcagctaag caaatagatg ataaatttat agttaaaaat tatttaataa tatcagcatg 45300
ttttttaatg gtagctatat tattatttaa tctatttcct tcattgatat ttaatcagga 45360
agtccactat agatatatag aaaaaatага tatgcttgtt acaagaatgg attttggact 45420
tggtaatcca aatagtgtat attatcatat ggttactata tatgcagcat atatattctt 45480
aagatataaa gattataata aatgggatag aataatttta tttggttctg catttttgt 45540
ttatcaaacc acatatagta gaactggatt tttcactata ttagcaggat taatatttgt 45600
agaaataata agatggatag atataaaaaa aataaaagga ctaccaatgc tcctaaagac 45660
attaccgata atacttactt tatttagtgt aattattggt acagtatttg ataaaagcac 45720
tttattaaat agattattag caagtagacc taagttctgg catgtatact tggctgaaaa 45780
ggggaacttt ctaaaacctt ttggaaattc atatagtcct gctataaaag caactaatcc 45840
attagatagt tcttatgtat atataattag tgtactagga gtagttgctt gtatattatt 45900
tatgtatttg atgtataaag gaatagaaag ctttatagaa aaagataaaa aagcatattt 45960
agtagcttta tttatatttt tactatactc ttttgcagaa aatttattat tagaagcttc 46020
atttagtttt gcagtagtat tactaataaa agaagtgata ttaaatgata aaagggaaat 46080
agatttgtgg aagatgaaaa gtaggaggta gtaatgttaa tatcattaat aatgccaaca 46140
ttaaatagat atgatgatat atacttactt atggacagct tagaaaatca aacttacaaa 46200
aactttgaac ttatagtagt agaccaaaat gataatagca aagttaaaga aatagttgat 46260
aaatacattg ataagttaga tataaagtat ataaaaagct ctaaaaaagg attgagttat 46320
aacagaaatg taggcataga caatgctgta gggcaaatta taggtttccc tgatgatgat 46380
tgtgtctacg aaaatgatac attagaaaaa gttataaatt tctttaataa aaataaagat 46440
tataaaatat acagttgcaa aactatggat tcaaataaag ttgatgcatt taagaaaatg 46500
tatgatggaa cttgtgatat cacaagttcc aacgttttgg atactataac atccataact 46560
ttttttattg attttgaagg taaagactat acaagatttg atgagaaatt gggtgttggt 46620
ggagagtttg gtgctggtga agagatagat tatgttttaa acttattgag tttaggtttt 46680
aaaggaaaat actttggaaa tgatataatt taccatcctg caaaaaaaca ttctaaatct 46740
aaagaaaagt atcaaaaaga ttataactat ggaagaggct ttggagcgct ttgcaaaaag 46800
gaaatagtct atagaaaaaa ttataagttt gcaaaagtta tggtatcaaa acttgtaaga 46860
aatataggag gtcttatatt gagttccaat agagattatc atagtgccac tataaaagga 46920
agaattaatg gatttagaca gtacaagtta taatgaggat ggtgtttatg gacaaggtta 46980
aatttttaat aaatatgctt ttggcttttt ttatatatcc atttaataag cataagttta 47040
atggtagaga tatctggctt ataggtggac attcagggga tatatacaat gataattcaa 47100
agtttttta tgagtatatg ctaaaagaac ataatgatgt agaaacttat tgggtagtaa 47160
ataaagacag taaagtattt gataaaattc caggcaaaaa attaataaga ggaagtgtag 47220
aaaactatct gtattattac aattcaaaag ctatagtatt ttcacatgca ccatcagcag 47280
atatagcacc atacaatttt gctgtacctg ttcttaatta ttttcataaa aaaacgataa 47340
aagtgttttt aaatcatggg actataagtt ttaaaaaaag aaaacctatg aacaaaaagt 47400
tcaaaaatat tatagataac ctatataaaa gctataatat agttacagca agttctgaat 47460
ttgaaagaaa tgttatggta aatgattggg gaatgttaga tgactcagtt tatataattg 47520
gaaatgcaag atatgacaat ttaccaacaa atgaagttgc acaaactcga gatatattat 47580
atactcctac atggagagat tggataaaat ttagttcagg aaaatttacg gatacagatt 47640
attttaaaaa tataatgaat tttttaaatg atgacaaatt aaataaaata ttagatgaga 47700
aagatataaa tgtaaaaatt tatatgcatc atcttatgca tgagtttata gatgatataa 47760
aagaaaatat aacaggaaaa cgaattgtgt ttttagataa aggggtaact ctagcaaatg 47820
aaatcagaaa atcagctgca aatataactg attattcaag tgtagccata gacttttttgt 47880
atatgaatag acctattttg ttttaccaat tcgatttgga tgaatatatg gaaaaagttg 47940

attcatatat agatttaaaa agtgaaatgt ttgggtcttt ggcttataat aatgatgaag 48000
ccgtgaataa acttattgat attattgaaa ataattttga ggttatggat aaccaaaaaa 48060
atgaaagaaa taagtttttc agatataatg ataataagaa ctgcaagagg atatatgatt 48120
gtgtattaag taaaattaaa taaatatttt ataaatgatg aggaaggtaa tatgaagaag 48180
aatttagtat ctataattac tcccatgtat aattctgaaa aatttattga agcaaccata 48240
aaatcagtat taaaccaaac ttatcaagaa tgggaaatgt taattattga tgattgctca 48300
acagataata gtcctaatat agtcaaatct tatatgcaac aggatagtag aataaaatgt 48360
ataaagactg agactaataa gggtgtctct aatgctagaa atttagcact aagtaaggca 48420
acaggacaat ttatagcttt tttangtagt gatgaccaat ggaatagtag taagttagaa 48480
aaacaagtaa attttatgtt agaaaatgac tatgtaattt catttacttc atatgaactg 48540
atggatgaaa a 48551

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 01/01280

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: C12N 15/63, A61K 39/08
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: A61K, C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, WPI DATA, PAJ, EPO-INTERNAL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 9519371 A2 (SOLVAY), 20 July 1995 (20.07.95), page 12, line 1 - line 9, claim 12, abstract --	1-33
Y	Microbial Pathogenesis, Volume 28, 2000, Marina Cerquetti et al, "Characterization of surface layer proteins from different Clostridium difficile clinical isolates" page 363 - page 372 --	1-33
P,Y	Molecular Microbiology, Volume 40, No 5, 2001, Emanuela Calabi et al, "Molecular characterization of the surface layer proteins from Clostridium difficile" page 1187 - page 1199 --	1-33

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 Sept 2001	02 -10- 2001
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Carolina Gómez Lagerlöf/BS Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 01/01280

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,Y	Infection and Immunity, Volume 69, No 5, May 2001, Tuomo Karjalainen et al, "Molecular and Genomic Analysis of Genes Encoding Surface Anchored Proteins from Clostridium difficile" page 3442 - page 3446	1-33

--

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE01/01280

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 30-32
because they relate to subject matter not required to be searched by this Authority, namely:

see next sheet

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Claims 30-32 relate to methods of treatment of the human or animal body by surgery or by therapy (PCT Rule 39.1(iv)). Nevertheless, a search has been executed for these claims. The search has been based on the gene expression cassette according to claims 1-10.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/SE 01/01280

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9519371 A2	20/07/95	AU 6380194 A	24/10/94
		EP 0691845 A	17/01/96
		EP 0738278 A	23/10/96
		GB 2291594 A,B	31/01/96
		GB 9400650 D	00/00/00
		GB 9519866 D	00/00/00
		JP 8508474 T	10/09/96
		JP 9508012 T	19/08/97
		US 5874267 A	23/02/99
		US 6028098 A	22/02/00